Selected Topics in Scattering Theory: From Chaos to Resonance

Selected Topics in Scattering Theory: From Chaos to Resonance

A dissertation presented
by
Bo Liu
to
The Department of Physics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Physics

Harvard University
Cambridge, Massachusetts
January 2015

Dissertation Advisor: Eric J. Heller Bo Liu

Selected Topics in Scattering Theory: From Chaos to Resonance

Abstract

Scattering problem is one of the most fundamental problems in physics, spanning almost all areas of physics. In this dissertation, we focus on scattering theory in two types of systems: two dimensional electron scattering in the presence of a random potential and light scattering by metallic nanoparticles.

The first scattering problem we study is electron branched flow. In this system, electrons are confined to move in two dimensions while a smoothly changing weak random potential deflects their trajectories, resulting in the so-called branched flow. A semiclassical theory based on ray tracing was developed to explain all the observed features of branched flow. However, this semiclassical theory was challenged by the result of a more recent experiment, which claims to have uncovered ”unexpected features of branched flow”. We show how these features can actually be explained by the semiclassical theory.

Besides electron scattering, we also investigate light scattering by metallic nanoparticles. In this case, we study the multiple scattering effect in the plasmon dimer system and show that one can use these metallic nanoparticles to put the incoming electromagnetic fields into different shapes by solving an inverse scattering problem.

Contents

Listing of figures

Dedicated to my parents, Qiaozhen Kuang and Siping Liu.

Acknowledgments

First of all, I want to thank my dissertation advisor, Prof. Eric Heller, without whom this dissertation would not be possible. During my years in his group, Rick has been nothing but supportive. Rick gives me a great amount of freedom to pursue what I find interesting and trains me to be an independent researcher. I am always captured by Rick's deep understanding of physical problems and how he could see through the math and comes up with an intuitive explanation. Rick taught me to think as an independent researcher and beyond the math, for which I am deeply grateful.

Secondly, I want to thank Prof. Bertrand Halperin and Prof. Robert Westervelt for being on my committee. I would also like to thank Dr. Mario Borunda for guiding me at the early stage of my research. Harvard Physics Department is an amazing place with many great people and I want to thank everyone who has helped me during the past five years, especially Judy Morrison, Lisa Cacciabaudo, Carol Davis, Dayle Maynard, Maggie McFee and William Walker.

Last but not least, I want to thank my parents, Qiaozhen Kuang and Siping Liu, for their unfailing support and unconditional love.

0

Introduction

Scattering theory is one of the most fundamental theories in physics and many problems can be formulated as scattering problems. In high energy physics, physicists are interested in exploring what would happen when particles with extremely high energy collide with each other. This type of problem has important implications for the foundations of theoretical physics and these high energy colliders are common test ground for new theory or for searching new fundamental particles. In condensed matter physics and biophysics, scattering problems are also ubiquitous. Many experimental probing techniques are based on scattering theory, including

electron loss spectroscopy, Raman spectroscopy and Scanning Tunneling Microscopy(STM).

Generally speaking, scattering problems can be divided into two categories: elastic scattering where energy is conserved and inelastic scattering where energy is not conserved. It is worth mentioning that if one takes into account all possible degrees of freedom, energy is always conserved. In this sense, all scattering events are elastic. The difference between elastic scattering and inelastic scattering lies in whether exchange of energy between different degrees of freedom is possible. In this dissertation, we will mainly focus on elastic scattering and we study two types of systems: electron scattering in the presence of random potential and light scattering by metallic nanoparticles.

The first interesting scattering problem we study is called branched flow[1,2], which was first discovered by Prof. Bob Westervelt in an amazing experiment on two dimensional electron gases(2DEGs) found at the interface of semiconductor heterostructures. In this system, electrons are cooled to extremely low temperature (4.2 K), so their behavior is dictated by quantum mechanics. Moreover, as the name suggests, the electrons lose one degree of freedom due to a lateral confinement and can only move in two dimensions. Such 2DEGs is also a common playground for studying Integer and Fractional Quantum Hall Effect, which leads to two Nobel Prizes in Physics. In his experiment, Prof. Westervelt was interested in seeing how electrons would be scattered by a weak random potential due to the donor atoms and the unavoidable impurities. The donor atoms contribute to a weakly correlated random potential that weakly backscatters the electrons, while the potential generated by the impurities can be both sharp and strong backscatters. Overall, the random potential is weak and the standard deviation of the random potential is only about 8% Fermi energy. Moreover, the sample size is also smaller than the mean free path of the electrons, so most electron shall be able to fly through the random potential without much deflection. However, as observed in the experiment, such weak random potential can nevertheless cast a strong signature on the electron paths.

As we will explain more in the following chapters, there are at least three very interesting observations in the experiment. The first one is the main topic of the experiment, branched flow. When flying through this random potential, electrons tend to move together along a small number of narrow branches rather than flowing uniformly through the sample. Secondly, persistent fringing is observed in the experiment, which goes well beyond the thermal length. The thermal length describes how far away interference shall be masked by the energy uncertainty due to finite temperature. In the experiment, the thermal length is about five electron wavelengths, while fringing goes beyond seventy wavelengths. The last surprising observation is the fine spatial resolution in the experiment. In the experiment, the measurement is done by holding a scanning gate tip on top of the sample and recording the change in measured conductance as a function of the position of the tip. The tip has a distance around two wavelengths above the 2DEGs. Even for a perfect point tip, this would translate into a scattering potential with width on the order of a few wavelengths. However, the spatial resolution in the experiment is better than half a wavelength. This fine spatial resolution can be explained by the so called ”glint effect”.

Following the discovery of branched flow, a semiclassical theory has been developed by Prof. Eric Heller in collaboration with Prof. Westervelt to explain its formation. According to this semiclassical theory, the formation of branches is explained by a ray tracing approach where caustics and stability regions in phase space are cited as the reasons for branch formation. Latter on, Prof. Heller has also extended this ray tracing approach to explain the formation of freak waves in oceans[8] and similar experiment has also been done in microwave cavities[4].

The semiclassical theory was widely applied until it was challenged by a new experiment by a group of researchers from Stanford University. In this experiment, the researchers reasoned that if the semiclassical theory is correct, the classical dynamics of electrons moving in a ran-

dom potential is chaotic, meaning that a small perturbation in initial condition would lead to large changes in the subsequent electron flow patterns. In the popular culture, this is known as the ”butterfly effect”, which says that the flipping of a butterfly’s wings in New Mexico could potentially lead to a hurricane in a far region like China.

In their experiment, the Stanford researchers first measured the branched flow pattern with one set of initial conditions and they were then able to shift the initial conditions by a large distance such that the two initial conditions have no overlap to begin with. They then measured the shifted branched flow pattern over the same random potential. Much to their surprise, the two sets of branched flow patterns look almost identical far away from the injection regions. This stability seems to contradict the semiclassical theory given the chaotic nature of the semiclassical dynamics and thus is termed ”the unexpected features of branched flow”. In this dissertation, we will show why such stability arises and how it could be explained within the semiclassical interpretation.

In latter chapters, we switch gear to study a different type of system, namely Localized Surface Plasmon. Plasmons are the collective excitation of electrons in metallic structures. Depending on the different boundary conditions, plasmons can be categorized into three types: volume plasmons, Surface Plasmon Polaritons and Localized Surface Plasmons. Volume plasmons are probably most familiar to the condensed matter physics community as one of the most well known examples of collective excitation. Volume plasmon is essentially the collective movement of electrons in an infinite metal where boundary conditions do not matter. It is essentially a bulk state and has routinely been observed using electron energy loss spectroscopy.

Surface Plasmon Polaritons and Localized Surface Plasmons are more well known in the nanophotonics community. They are essentially volume plasmons in the presence of different boundary conditions. Surface Plasmon Polaritons arise at the interface of a semi-infinite

dielectric material and a metallic structure. It is essentially the coupling between light and the electrons on the interface. Surface Plasmon Polaritons can not in generally be directly excited by laser beams and require near field excitation. Localized Surface Plasmon, on the contrary, can directly couple to exciting light and are the result of the coupling of light and the volume plasmon subject to the boundary conditions set forth by the shape of a nanoparticle. In its essence, the Localized Surface Plasmons can be thought as a resonant electric dipole moment. In this dissertation, we will mainly focus on Localized Surface Plasmons.

For most types of scatterers, its scattering cross section is roughly on the same order of magnitude as its physical cross section. This is especially true for classical hard wall scatterers. One of the surprising consequence of quantum mechanics is that for hard-wall s-wave scatterers, its scattering cross section is twice its physical cross section. This gets more interesting when the scatterers have internal degrees of freedom. For instance, the magnetic resonance imaging(MRI) is based on the scattering of electromagnetic wave by the nucleus of atoms. In that case, the internal degree of freedom is due to the nuclear spin, which has different energy levels in the presence of a magnetic field. In general, when the energy of the object being scattered matches the energy difference between the internal degree of freedom, one should expect resonance to take place. One interesting property of resonant scatterers is that its scattering cross section as λ^2, where λ is the wavelength of the object being scattered, usually electrons or photons, under the resonant condition. For small gold nanoparticles, the resonance wavelength is around 540nm, while the radius of the gold nanoparticle can be as small as 5nm. As a result, the scattering cross section of such gold nanoparticles can be four orders of magnitude larger than its physical cross section. If one places two such gold nanoparticles well within one wavelength, but outside of each other's physical range, we can expect to see a regime where multiple scattering has the main effect. In quantum scattering theory, if one places two resonant s-wave scatterers under such conditions, one should expect to observe

a symmetric mode with increased damping rate and an antisymmetric mode with reduced damping rate. This is known as the quantum proximity resonance[37]. For gold nanoparticles, the physics is similar, but also different due to the existence of resistive loss. In this case, we show that one can get an antisymmetric mode with reduced damping instead.

For the last part of this dissertation, we solve an inverse scattering problem involving these gold nanoparticles and show how this approach can be used to control the shape of electromagnetic fields.

In recent years, tremendous interest has been put towards controlling electromagnetic fields. In 2000, Sir John Pendry proposed to use left-hand materials to build a perfect lens[76], which could in theory eliminate the diffraction limits set upon traditional lens. Of course, left-hand materials with negative refractive index does not exist in nature. However, with the successful fabrication of meta- materials, such material with negative refractive index was finally demonstrated experimentally in 2001[28].

Later on, Prof. Pendry moved on to propose another genius idea that again spurred tremendous interest in the nanophotonics community, namely transformation optics[73]. In the original paper, Prof. Pendry showed that if one can accurately engineer the electromagnetic properties of a material on the subwavelength, one can then shape electromagnetic fields at will. As a proof of concept, he showed that one can realize an invisibility cloak by using a shielding shell made of metamaterials.

The key to transformation optics is to use materials with heterogeneous electromagnetic properties in a controlled manner. Prof. Pendry demonstrated a mathematically equivalence between shaping electromagnetic field and varying the electromagnetic response of the material in a heterogenous manner. According to transformation optics, one can achieve any electromagnetic pattern if one has perfect control over the electromagnetic properties of the constituent unit of the materials. However, this also implies that sometimes one needs to change

the permeability and permittivity of a material rapidly on the sub-wavelength scale and also achieve certain extreme values at the same time in order to achieve certain functionality, such as cloaking.

In the last part of this dissertation, we propose a different approach to achieve the control of electromagnetic field on the sub-wavelength scale without the need to engineer a material's electromagnetic properties. Instead, the control over the electromagnetic field is achieved by controlling the positions of the metallic nanoparticles. We reformulate this problem of controlling the electromagnetic fields as an optimization problem, where the free variables to adjust are the positions of the metallic nanoparticles. With only 80 gold nanoparticles, this combinatorial optimization problem has a configuration space of size of 10^{235}, which is even larger than the total number of atoms in the observable universe. This enormous configuration space presents both challenge and promise for finding a solution for the optimization problem. We solve this optimization problem by employing a heuristic algorithm known as the simulated annealing and we are able to put the electromagnetic fields into the shapes of the alphabetical letter ”H”.

1

Discovery of Branched Flow

Branching is a universal phenomenon of wave propagation in a weakly correlated random medium. It is observed in 2DEGs with wavelength on the scale of nanometers [1,2], in quasi-two-dimensional resonator with microwave [4] and used to study sound propagation in oceans with megameter length scales [5]. It has significant influence on electron transport in 2DEGs [6,7] and is found to be implicated in the formation of freak waves in oceans [8].

1.1 First Discovery of Branched Flow

Branched flow was first observed in an experiment by Prof. Westervelt[1,2] on two dimensional electron gases(2DEGs) in GaAs heterostructure. The experimentally observed branched flow pattern[1] is shown in Fig. 1.1 below.

There are at least three surprising features of this observed pattern. The first surprise is of course the branched flow pattern itself. It was estimated that the random potential due to charged donor atoms and impurities is only about 8% the energy of the electrons. If one naively looks at this number and thinks in terms of a perturbation theory, one wouldn't expect such a week random potential to lead to such significant changes in the electron flow pattern.

The second surprising observation is the persistent fringing observed here. As one can see from the figure, all the branches are decorated with fringes, which persist for a few micrometers. However, a simple calculation based on the thermal length[10] at this experimental temperature would indicate that the fringes should be wiped out by thermal averaging within a micron or so.

The final surprise is related to the spatial resolution of the experiment. The branched flow pattern is observed using a charged tip held about 60nm above the 2DEGs. This indicates that the tip has to be at least more than 60nm in width. However, the spatial resolution in the experimental result is on the order of 10nm or so, which is smaller than the width of the probing potential as generated by the measuring tip.

All these surprises can be explained within a thermal wave packet approach as developed in Ref.[10].

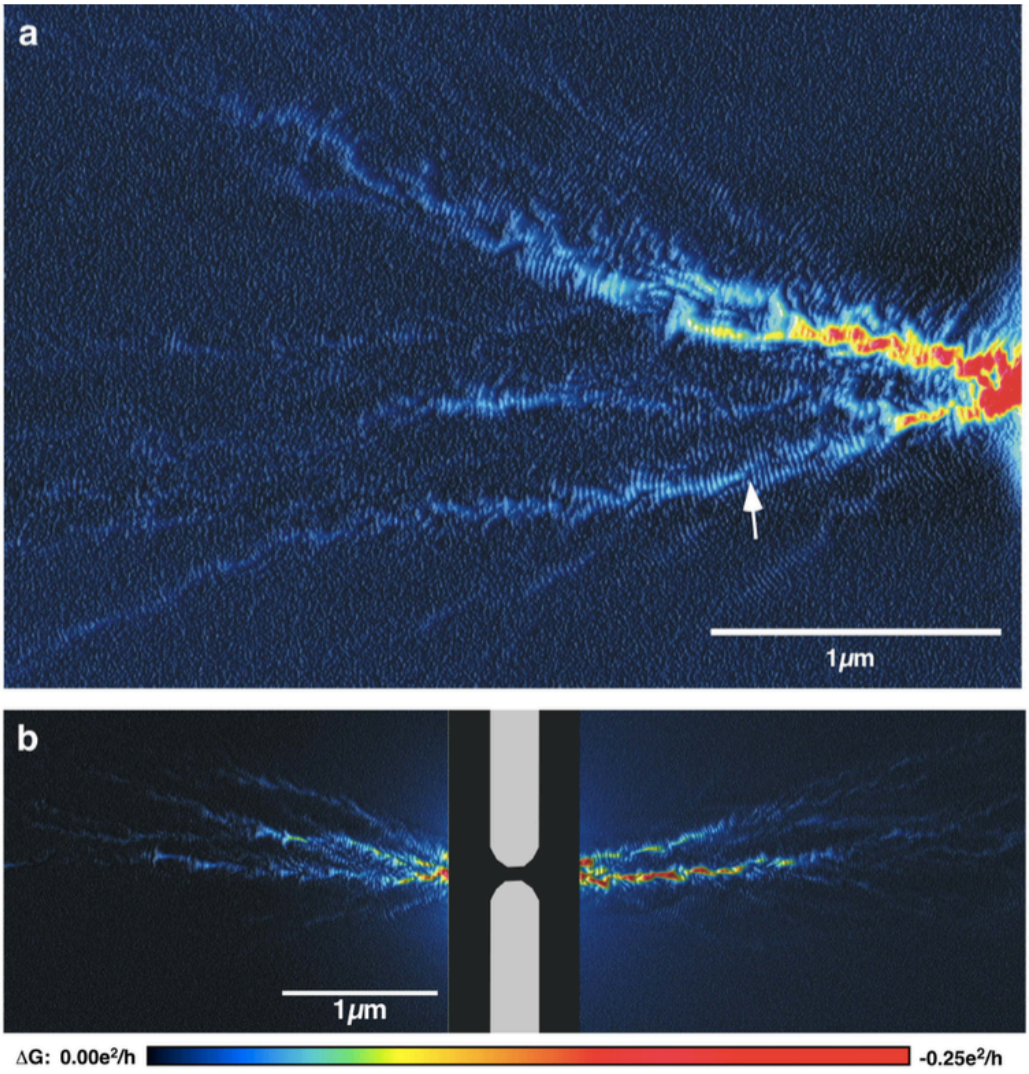

Figure 1.1: Experimentally observed branched flow. This figure is taken from Ref. 2. In this figure, we see the experimentally observed branched flow pattern. The two patterns in (b) correspond to two different random potential. This figure is obtained with the copyright permission from Nature Publishing Group.

1.2 Experimental Setup

In Fig.1.2, we show the experimental setup used to probe branched flow. In the experiment, a Quantum Point Contact(QPC)(Blue regions) is formed by fabricating metallic gates on top of the sample. The QPC is negatively charged, so it behaves just like a waveguide and for this specific waveguide, usually only one or two modes will be open for conduction in the experiment. This QPC is used as the injector for the electrons.

The measurable in the experiment is the two terminal conductance. In the ideal case where one mode is open in the QPC, the measured conductance should approach the ideal value

$$G = \frac{2e^2}{h}. \tag{1.1}$$

However, backscattering due to the charged donor atoms and impurities will in general make the measured conductance smaller than this ideal value.

In the experiment, two sources of backscattering exist and they have different strengths in backscattering the electrons. The first source of backscattering is due to the random potential generated by the charged donor atoms. The donor atoms are usually separated from the 2DEGs by a spacer layer, so they generate a weakly correlated random potential that do not strongly backscatter the electrons. The second source of backscattering is due to the unavoidable impurities introduced in sample preparation. These impurities can get very close to the 2DEGs and thus can become very sharp and strong backscatterers. The weakly correlated random potential is in general believed to be responsible for the formation of branched flow, while the strong backscattering impurities are responsible for the persistent fringing.

In the experiment, an additional source of backscattering is introduced to help decide the electron flow pattern. A metallic tip is held on top of the sample. The metallic tip is negatively charged, so it creates a potential barrier for all the electron flow impinging on it. This

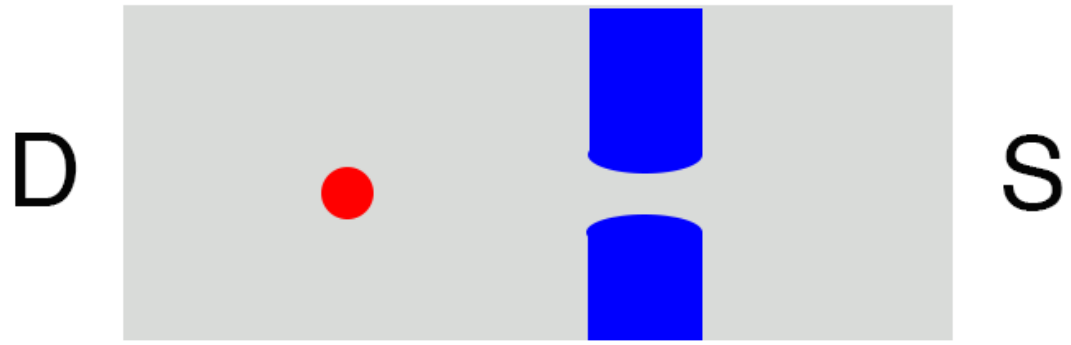

Figure 1.2: Experimental setup to observe branched flow.This figure shows the experimental setup used to probe branched flow. The measurable in the experiment is the two terminal conductance and a charged metallic tip(red dot) is held on top of the 2DEGs, which backscatters the electrons flow through the regions underneath. When the tip is held on top of region with high electron flux, more electron flux will be backscattered, leading to a larger change in the measured conductance. On the other hand, if the tip is placed on top of a region with relatively low electron flux passing through, it leads to a smaller reduction in the measured conductance. Thus, by moving the tip across the sample and recording the change in conductance, one can get a pretty good picture of how electrons move though the sample.

potential barrier will backscatter the electrons moving through it. The more electron flow moves through that region, the more electron flow will be backscattered towards the QPC. Thus, if the tip is held on top of a region where a large electron flux is moving through, it will lead to a large reduction in the measured conductance and if it is held on top of a region where almost no electron flux passes through, it creates a small conductance reduction. Thus, by moving the tip across the two dimensional sample and recording the changes in measured conductance, one can get a good picture of how electrons move through the sample. This essentially constitutes the results reported in Fig.1.1.

1.3 Formation of Branched Flow

Since its discovery, classical theory has been put forward to explain the formation of branches. For instance, in the original published paper, Prof. Westervelt and Prof. Heller reported a classical branched flow pattern simulation that closely resembles the quantum flow simulation. In

the classical simulations, electrons are propagated according to Newton's equations and the initial conditions for the electrons are chosen to populate the whole classical phase space.

According to the classical theory [14,15], two reasons can lead to branched flow: caustics in phase space and stable regions in phase space.

Caustics forms after electrons pass through a potential dip, which focuses parallel electron trajectories. Immediately after the focal points, one caustics forms where electrons travel along with high density.

One theory of the origin of branched flow is called the "kick and drift" model [15], which is closely based on the idea of caustics. According to this "kick and drift" model, the effect of the random potential is to give each electron a random kick in the phase space, after which electrons are allowed to drift freely. Then another kick and another drift. This kick and drift process gets repeated many times and in the process, random caustics appear, which accounts for the formation of branches.

For electrons moving in two dimensions, the phase space should in general be four dimensions, with two dimensions for positions and two dimensions for momentum. Energy conservations reduce this four dimensional phase space to three dimensions. A further reduction in dimension can be achieved by the so-called quasi-one dimensional approximation. For the branched flow system, the random potential is very weak ($8\%E_F$). Since most of the electron energy is focused in the longitudinal direction(direction of propagation), the quasi-one dimensional approximation is essentially saying that one can ignore the changes in the momentum in the longitudinal direction and assume it is a constant. With this assumption, the original three dimensional phase space becomes two dimensional.

For a detailed analysis of the distribution of caustics in a random potential, we refer the readers to Ref. [14].

Besides caustics in phase space, another mechanism could also lead to the formation of

branched flow, which is known as stable regions in phase space[11,15]. In a nutshell, stable regions in phase space form by chance. This idea is best illustrated in Fig.1.3.

In Fig.1.3(a), the stuffed circles correspond to regions in phase space. Those are the initial conditions taken by the electrons. In Fig.1.3(b), electrons are allowed to propagate in the random potential for five correlation lengths. Different circles get distorted by different amount. For the regions circled by the red line, the initial stuffed circle was less distorted when compared with the others and these are possible candidate for branches. In short, stable regions in phase space correspond to regions that do not get distorted much when compared with the rest. These regions, when projected into the real space, correspond to places where one observes high electron flux density.

One way to understand those stable regions in phase space is like this: one fundamental property of the phase space structure is that the total electron counts do not change. As long as the total area being occupied do not expand exponential, as it shouldn't in this case due to the fact that the sample size is smaller than the mean free path, when certain regions' electron density is diminished, this decrease in electron density has to be compensated with a higher electron density somewhere else. These more densely populated regions are stable regions.

One way to characterize the stable regions is called the rarefaction exponent[11,15], which is a variant of the Lyapunov exponent. The definition of the rarefaction exponent is

$$r(t) = log|M_t\vec{d}|, \tag{1.2}$$

where M_t is the stability matrix as introduced below and $\vec{d}$ is a unit vector in a direction of interest in the phase space.

The stability matrix is defined on a per classical trajectory basis and it is essentially a mapping function from a small perturbation to the initial condition of a fixed classical electron trajectory to subsequent changes.

The mathematical definition is

$$\delta\vec{x}(t,\vec{x}(0)) = M_t(\vec{x}(0))\delta\vec{x}(0). \tag{1.3}$$

In the above definition, $\vec{x}$ is a vector in the four dimensional phase space, that is $\vec{x} = (x, y, p_x, p_y)$ and $\vec{x}(0)$ is an initial condition for a given classical electron trajectory. Suppose we give a small perturbation to the initial condition such that the perturbed initial condition is $\vec{x}(0) + \delta\vec{x}(0)$, then at a latter time t, the perturbed trajectory is defined relative to the unperturbed classical trajectory as $\vec{x}(t) + \delta\vec{x}(t,\vec{x}(0))$, where $\vec{x}(t)$ is the classical trajectory for the unperturbed initial condition and $\delta\vec{x}(t,\vec{x}(0))$ is how large the initial perturbation is magnified up till time t.

By this definition, we can write down each element of the stability matrix as

$$M_{i,j}(t) = \frac{\partial\delta x_i(t)}{\partial\delta x_j(0)}. \tag{1.4}$$

In its strict sense, each element of the stability matrix should be calculated numerically. One should keep track of two classical trajectories whose initial condition differing by a tiny amount and calculate their difference at each time. This is of course more exact, but clearly is also not very numerically efficient. One alternative method commonly used to characterize and calculate the stability matrix is by assuming local dynamics, that is by assuming that even though the classical dynamics is chaotic, two trajectories with infinitesimal difference in initial conditions will stay infinitesimally close to each other's neighborhood. Thus, one only need to keep track of one classical trajectory and calculate all the elements of this stability matrix based on that single classical trajectory. The detailed derivation is given in Ref.[11] and we shall only present the result here. The result is that one can derive a dynamic equation for

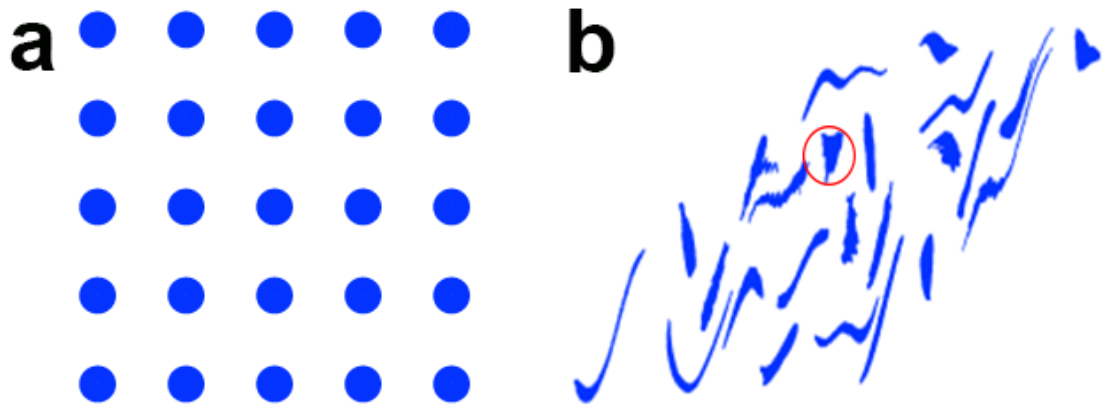

Figure 1.3: Stable regions in phase space. (a) The initial conditions in phase space for electrons. The unit circles are uniformly populated. The horizontal axis corresponds to the position in the horizontal direction and the vertical axis corresponds to the momentum in the same direction. (b) The structure in phase after moving five correlation lengths in the random potential. The region circled by the red line is what we call a stable region and is a candidate for the branches.

calculating the stability matrix:

$$\frac{dM_t(\vec{x}(0))}{dt} = K(t)M_t(\vec{x}(0)), \tag{1.5}$$

where

$$K(t) = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \left[\frac{\partial^2 H}{\partial x_i \partial x_j}\right], \tag{1.6}$$

and H is the classical Hamiltonian of the system.

Thus, in order to calculate the stability matrix, one only needs to keep track of one single classical trajectory. By choosing the different measures of convergence, one probably could even make this argument mathematically sound. However, in any realistic situation, only a stability matrix with small eigenvalues will have interesting physical meaning.

For more information on this rarefaction exponent, we refer the readers to Ref.[11] and Ref.[15].

It is also worth noting that in plotting Fig.1.3, we do not necessarily need to assume the quasi-one dimensional approximation as assumed before. Here the dimension reduction is

due to a different reason. As before, energy conservation reduces the four dimensional phase space to three dimensions. If we label this three dimensional phase space by three independent variables (x, y, p_x), then the further dimensionality reduction comes from fixing y. So what we essentially are plotting in Fig.1.3 is a slice of this three dimensional phase space.

Strictly speaking, the kick and drift model also does not require the quasi-one dimensional approximation. However, the quasi-one dimensional approximation could make the thinking process much easier. When one assumes the quasi-one dimensional approximation, the propagation time from y to $y + \delta y$ is the same for all electrons and the random potential is then essentially contributing to a momentum kick and subsequent shift to the electron's trajectories. If one takes into account the effect of the random potential onto the momentum in the longitudinal direction, the analysis is much harder since each electron will travel different amount of time from y to $y + \delta y$ and there might also be some cumulative effect involved as well.

The real advantage of this quasi-one dimensional approximation lies in how it could simplify analytic analysis. For more on this, we refer the readers to the materials presented in Ref.[14]. Also Dr. Jakob J. Metzger's PhD thesis includes most of the classical analysis of branched flow based on this approximation. If the readers are interested in learning more background on branched flow and its explanation, we also refer the readers to the PhD thesis by Dr. Scot Shaw[11] and Dr. Mark Topinka[18].

To sum up this chapter, branching is a universal phenomenon of wave propagation in a weakly correlated random medium. Since its discovery, much effort has been put towards developing a theory to explain its origin. One of the most promising theory is based on the classical dynamics of electrons in a random potential. According to the classical theory, branches form as a result of both caustics and stable regions in phase space.

2

Unexpected Features of Branched Flow

Branched flow was discovered in 2000 and much theoretical work and experimental work have been done in the following decade. The main contribution of this chapter is to explain the results of a more recent experiment[9]. A significant portion of the materials in this chapter has been published in Ref.[3].

2.1 What does the experiment say?

Since the main contribution of this thesis to branched flow is to explain the results of this experiment, it is worthwhile to spend a section explaining what is exactly done in this experiment.

There are two major results of this experiment done by Prof. David Goldhaber-Gordon's research group at Stanford University. First of all, it verifies an existing theory that says branched flow is caused by the random potential and fringing is caused by the unintentional impurities. Secondly, it unveils some "unexpected" features of branched flow which seem to contradict the existing semi-classical theory of branched flow. It is this second observation that is the main topic of this chapter and we will demonstrate how this unexpected feature can actually be explained within the existing theory.

2.1.1 Origin of Fringing

As mentioned previously, one of the most striking observations of the first experiment on branched flow is the persistent fringing beyond the so called thermal length. This observation was already explained theoretically using a thermal wave packet approach in Ref.[10], which we will also discuss in later sections.

According to the thermal wave packet approach, fringing occurs due to the interference of the backscattered wave from the tip and the backscattered wave from the unintentional impurities. A more detailed explanation will be presented in later sections, but for now, let's accept this as it is and show how this experiment validates this theory.

The reasoning behind the results of the experiment can be summarized as the following: if fringing is indeed caused by the interference between the backscattered wave from the scanning tip and the backscattered wave from the impurities, then if one is able to reduce the densities of the impurities, one should also be able to see suppressed fringing in the measured

branched flow pattern.

Thus what Prof. David Goldhaber-Gordon's group was able to achieve is to find very high quality samples with varying amount of impurities and compare the different fringing patterns across different samples.

The quantitative measure of the quality of a sample is the mobility or the mean free path of the electrons in the sample. The reason this is correlated with the density of the impurities is because impurities are the major source of backscattering in these samples. The weakly correlated random potential is not very effective in backscattering the electrons due to both their weak strength and their broad scope compared with the wavelength of the electrons. The impurities, on the other hand, can be both sharp and strong scatterers.

In the experiment, the mobility of the three samples differs by one order of magnitude and it is indeed observed that samples with high mobility tend to show suppressed fringing. Thus, this experimental observation agrees with the thermal wave packet approach.

2.1.2 Unexpected Stability of Branched Flow

Before we go into details of this part of the experimental results, it is better to first discuss a little bit more about the motivation and reasoning behind. This part of the experiment is essentially trying to test how chaotic the quantum branched flow pattern is or how sensitive it is in response to a large perturbation in initial conditions. If the semiclassical theory of branched flow is correct, the classical dynamics of electrons moving in a random potential is actually chaotic, meaning that a small change in initial conditions will be exponentially magnified over time. Thus, one might assume that if the classical theory can accurately explain the formation of branched flow, then the branched flow pattern will differ significantly from the original branched flow pattern if a large perturbation to the initial conditions is created.

To create a large change in initial conditions, the QPC or the injector of the electrons was

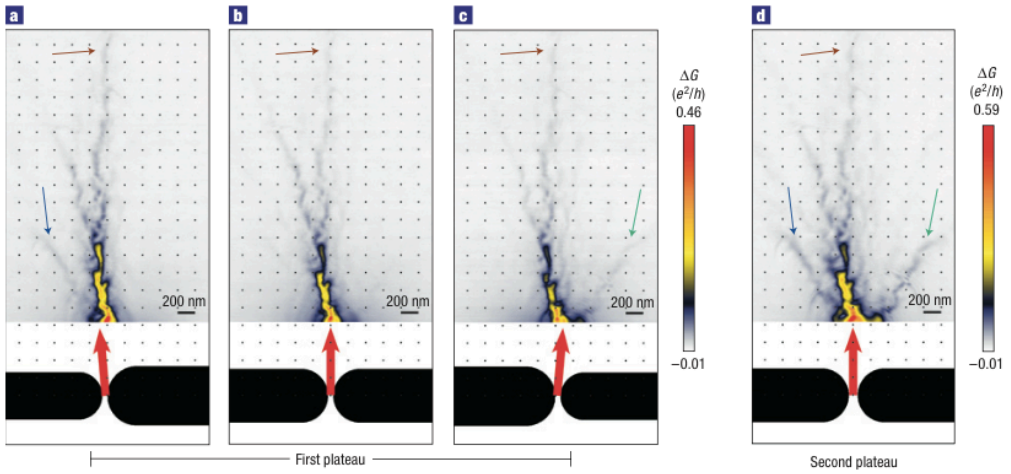

Figure 3 SGM images of branch stability in sample B. This figure shows the remarkable experimental stability of branched flow. The reference grid (200 nm) denotes the same physical location for precise comparison of electron flow between images. **a**–**c**, Images of electron flow with the QPC open to the first plateau of conductance and electrostatically shifted to the left, not shifted and shifted to the right, respectively. **d**, The second plateau of conductance with gates symmetrically energized. Black regions indicate schematically the depletion regions forming the QPC. The thick red arrows denote the initial average position and momentum of injected electrons. On changing the initial conditions of injected electrons, the location of branches does not change, but the current density flowing along each branch does. In all images, the central branches, one of which is labelled with a brown arrow, are visible. In **a** and **d** the left branches, labelled with a blue arrow, are visible. In **c** and **d** the right branches, labelled with a green arrow, are visible.

Figure 2.1: A figure take from Ref.[9] with its original captions.This figure is reproduced with the copyright permission from the Nature Publishing Group.

shifted by a large distance with regards to each other. The branched flow patterns were then measured over the same random potential with the shifted initial conditions. The results are shown in Fig.2.1. As we can see, the branches pointed to by the red arrow are present in all the four figures. Moreover, if one looks closely at the experimental results, the branches are actually at exactly the same location seventy correlation lengths away from the injection point. This lack of sensitivity even at long range was termed “the unexpected features of the branched flow”.

It is also worth noting that new branches do appear when the QPC is shifted, as pointed to by the blue and green arrow, and these new branches reappear when the second mode is open for conduction as shown in Fig.2.1(d).

These features were not satisfactorily explained in the original experimental paper and it was suspected that the observed stability is of pure quantum origin with no classical origin. As part of the proof of the claim, the authors also did both classical and quantum simulations that show how such stability can only be reproduced using quantum simulations.

However, we will show that the above claim is not true and we will also show how such stability could even be reproduced using only classical trajectories. The key difference between our simulations and the simulations done by the experimental group lies in how we handle the QPC.

In the following sections, we will develop a coherent picture that explains all these observed features.

2.2 Why was the observation surprising?

As mentioned before, we will show that the so called ”unexpected” features of branched flow could be explained using existing theory, so the question is then ”Why was the observation interesting in the first place?”

Of course, it is generally true that once we truly understand something, it is no longer surprising. But before anyone has an explanation and when a simple contradiction between the observation and the existing explanation exists, it is a surprise. So let’s briefly go over why the observation was a surprise and what the contradiction is.

In classically chaotic systems, small differences in initial conditions are exponentially magnified over time. However, it was observed experimentally that the (necessarily quantum) “branched flow” pattern of electron flux from a quantum point contact (QPC) traveling over a random background potential in two-dimensional electron gases(2DEGs) remains substantially invariant to large changes in initial conditions. The random potential in the experiment is classically chaotic and unstable to changes in initial conditions. So if one believes in the classical interpretation of branched flow, it seems likely that the branched flow pattern should also be substantially unstable to perturbations and no long-range stability shall be expected.

Before we move on to the explanations, we need to first to explain why the changes in initial conditions are indeed significantly both classically and quantum mechanically.

To create a large change in initial conditions, the QPC was shifted by about one correlation length of the underlying random potential, which is also roughly the width of the QPC. Classically, a one correlation length shift is indeed very significant for the chaotic dynamics, making the trajectories very different, as seen in the classical simulations of reference[9].

For quantum dynamics, if one launches two separate quantum wavepackets[10] through QPCs differing by this amount, the coherent overlap between the two initial wavepackets is estimated at less than five percent. As a Gedanken experiment, consider a pair of side-by-side

QPCs differing by a shift. This could not be in the experiment, which had only one QPC, which however was able to be shifted relative to the rest of the device and the branched flow imaged again. In the Gedanken experiment, suppose we put wavepacket A though one QPC and wavepacket B through the other. Can the coherent overlap between the initially nonoverlapping A and B wavepackets increase over time and distance from the QPC's? The answer is of course no, both classically (considered as overlap in phase space) and quantally. It is elementary to show that the coherent overlap must remain the same over time if the wavepackets are propagated under the same Hamiltonian. This is true whether or not disorder is present. So this change in initial conditions is also significant for the quantum dynamics.

Thus, the experimental result can be summarized as the following: the experimentally observed branched flow has a long range stability against changes in initial conditions even when the underlying classical dynamics is chaotic and exponentially unstable to perturbations. In the experiment, it is observed that some branches remain at almost exactly the same locations seventy correlation lengths away from the injection points, with the only observed difference being the relative strength of each branch. This lack of sensitivity even at long range was termed "the unexpected features of the branched flow".

2.3 Thermal wave packet approach

The thermal wave packet approach is a coherent theory developed to explain many of the observed features of branched flow[10]. It establishes a theoretical equivalence between the conductance in the system, which is essentially an eigenstate problem, and the solution of a time dependent dynamical problem.

In the experiments, the electron flow pattern is measured through the change in conductance as a function of the position of the scanning tip. Thus the starting point for the theory is the Landauer formula at finite temperature, which is given by

$$G_T = \frac{2e^2}{h} \int \frac{-\partial f_T(E)}{\partial E} Tr[\mathbf{t}(E)\mathbf{t}(E)^+] dE, \tag{2.1}$$

where G_T is the conductance at temperature T, $f_T(E)$ is the corresponding Fermi distribution and $\mathbf{t}(E)$ is the transmission matrix. In the presence of a QPC, there will only be a discrete number of modes open for transmission at any given energy E. Since we only care about the trace in (2.1), we can choose to work with the eigenmode of the QPC and add up the contribution from each mode separately, which is the underlying idea for the "thermal wavepacket" approach.

When only one mode is open in the QPC, equation (2.1) reduces to

$$G_T = \frac{2e^2}{h} \int \frac{-\partial f_T(E)}{\partial E} \tau_1(E) dE, \tag{2.2}$$

where $\tau_1(E)$ is the only nonzero eigenvalue of $\mathbf{t}(E)\mathbf{t}(E)^+$.

In this case, we can create an initial wavepacket of the following form

$$\psi_{1,T}(x,y,0) = \int dE \; exp(i\phi(E,y_0)) a_T(E) \psi_1(x,y,E) \quad , \tag{2.3}$$

where $\psi_1(x,y,E)$ is the scattering eigenstate with transmission probability $\tau_1(E)$, $\phi(E,y_0)$ is chosen such that it is a compact wavepacket centered at y_0 and $a_T(E)$ represents the energy distribution.

The time evolved thermal wavepacket is

$$\psi_{1,T}(x,y,t) = \int dE\, a_T(E)\, exp(-iEt/\hbar + i\phi(E))\, \psi_1(x,y,E). \tag{2.4}$$

Far to the right side of the injection point, the back-scattered part of the wave packet separates from the transmitted wave packet and we can replace $\psi_1(x,y,E)$ by $\mathbf{t}_1(E)\psi_1^{out}(x,y,E)$, where $\psi_1^{out}(x,y,E)$ is a unit flux outgoing wave and $\mathbf{t}_1(E)$ is the transmission coefficient.

Thus, the total transmitted flux is given by the following formula

$$F_T = \frac{4\pi e^2\hbar}{m} \int |a_T(E)|^2\, \tau_1(E) dE. \tag{2.5}$$

If we choose $a_T(E) = \hbar^{-1}\sqrt{-\frac{m}{2\pi}\frac{\partial f_T(E)}{\partial E}}$, we would get $F_T = G_T$, which shows that one can get all the information about the experimental conductance map by propagating this wavepacket through the scattering region.

For any given point, the time dependence of the flux can be calculated by applying the following flux operator to the wavepacket:

$$\hat{f}_s = \frac{1}{2}[\hat{n}\cdot\hat{v}\delta(\vec{r}-\vec{r}_s) + \delta(\vec{r}-\vec{r}_s)\hat{n}\cdot\hat{v}], \tag{2.6}$$

where $\hat{n}$ is the direction of interest, which is chosen to be the y direction in the simulations. $\vec{r}_s$ is the point of interest and $\hat{v} = -\frac{i\hbar}{m}\nabla$ is the velocity operator. For our quantum simulations, we report the total flux through a given point, which is $\int \psi_{1,T}^*(x,y,t)\hat{f}_s\psi_{1,T}(x,y,t)dtdxdy$.

When the second transmitting mode of the QPC is also open, we need to propagate a second

wavepacket separately to include its contribution. The total quantum flux will be an incoherent sum of the flux from both wavepackets. This is so because equation (2.1) only involves the trace of the scattering matrix.

Just as what we do for the first mode, we can prepare a second initial wavepacket of the following form:

$$\psi_{2,T}(x,y,0) = \int dE\, exp(i\phi'(E,y_0))a_T(E)\psi_2(x,y,E) \quad , \tag{2.7}$$

where $\psi_2(x,y,E)$ is the scattering eigenstate corresponding to the second eigenmode of the QPC , $\phi'(E,y_0)$ is chosen such that it is a compact wavepacket centered at y_0 and $a_T(E) = \hbar^{-1}\sqrt{-\frac{m}{2\pi}\frac{\partial f_T(E)}{\partial E}}$. To reproduce the experimental results, one just needs to add up the contributions from $\psi_{1,T}(x,y,0)$ and $\psi_{2,T}(x,y,0)$ incoherently.

This approach is of course not just limited to the first two modes in the QPC and in theory could be used to solve the electron flux pattern at any energy. However, for each mode, one needs to prepare a separate initial wave packet for each open mode in the QPC and then add up their contributions incoherently.

2.3.1 Thermal length

The width of the thermal wave packet is decided by the amplitude of each eigenstate at the given energy, which is

$$w(E) = \sqrt{-\frac{\partial f_T(E)}{\partial E}} = (4k_BT)^{-1/2}sech[(E-E_F)/2k_BT]. \tag{2.8}$$

One simplification comes in handy when we realize that the experiment is performed at extremely low temperature and the energy fluctuation is small. In this particular system, the energy of the electron is mostly distributed to its kinetic energy. Combined with the fact of small

energy fluctuations, this implies a very narrow wave packet in the momentum space.

With this intuition in mind, we can approximate the above formula by a Gaussian weight as[11]

$$\begin{aligned} -f'(E) &= \left[1 + e^{(E-E_F)/kT}\right]^{-2} \frac{1}{kT} e^{(E-E_F)/kT} \\ &\approx \frac{1}{4kT} e^{-(k-k_F)^2 \ell_T^2}. \end{aligned} \tag{2.9}$$

The standard deviation of the Gaussian in Eq.(2.9) was chosen to match both the value of $-f'(E)$ at $E = E_F$ and its approximate width while making sure that it is properly normalized[11]. ℓ_T is the thermal length, and it is given by $\ell_T = \hbar^2 k_F \pi^{1/2} / 4mkT$.

This thermal length decides the spatial scope of the thermal wave packet and clearly, the lower the temperature, the larger the thermal length. At 1.7K, ℓ_T =0.4 μm and 0.16 μm at 4.2K in a sample with $E_F = 0.016eV$ and $v_F = 2.86 \times 10^5$ m/s[10]. When compared with the wavelength of the electron, the thermal length divided by the wavelength is $\ell_F / \lambda_F = E_F / \pi^2 k_B T \sim 5$ in this example.

For numerical simulations, this Gaussian approximation is not necessary and the exact initial wave packet can be prepared using Fast Fourier Transform. However, this Gaussian approximation can be very useful in developing an intuitive understanding of the physical picture of the underlying physical process.

2.4 Fringing

As promised before, we will now use the thermal wavepacket picture to explain the persistent fringing. Nearly all the explanation given in this subsection is nothing more than a rephrasing of what was said in Ref.[10] and we strongly recommend the interested readers to read that paper and get a more comprehensive understanding of the physics.

We can imagine that the thermal wave packet is launched from one side of the QPC and as it emerges from the QPC, it encounters both the random potential due to the donor atoms and the sharp potential due to the impurities. The random potential due to the donors can only deflect the wave packet's initial trajectory, but the impurities can send part of the thermal wave packet back towards the QPC. The samples used in the experiment are usually of high mobility/quality such that the density of impurities is low and we can assume them as isolated scatterers not worrying about multiple scattering effect.The charged metallic tip used to probe the electron flux can also backscatter part of the thermal wave packet towards the QPC and if the backscattered wave from the impurities and that from the tip return to the QPC at the same time, they could interfere and it leads to the observed fringing.

For the two backscattered wave to return to the QPC at the same time, the impurities have to be within half a thermal length of the charged tip. In the experiment, only the position of the charged tip is changing, so we can assume that the backscattered wave from the charged impurities that can interfere with the backscattered wave from the charged tip is a fixed background as one moved the charged by a small distance.

The width of the thermal wavepacket at a typical experimental condition is about five wavelengths. As a result, the main effect of moving the tip by a small distance is to cause a phase shift of its amplitude. However, if one moves the thermal wave packet or the tip by a distance larger than the width of the thermal wave packet, the backscattered wave from the tip is then interfering with the backscattered wave from a completely new set of impurities. Thus, the

fine structure of the fringes can be different when radially separated. This physical picture also explains why the fringes seen in the experiment is approximately radially oriented towards the QPC. If one moves the tip along a trajectory of constant radius with regards to the QPC, the backscattered wave from the tip will have approximately the same phase. Given that the backscattered wave from the impurities does not vary significantly, small movement of the tip along the arc will lead to approximately the same interference amplitude.

This thermal wave-packet picture also explains why the thermal averaging does not destroy the fringes. Thermal averaging destroys interference by adding up waves with different $\vec{k}$, which have different nodal structures. When one adds up different waves with different nodal structures, it could potential destroy each other's nodal structure, leading to the destruction of interference patterns. In this case, the thermal wave-packet moves across the sample and waves get backscattered towards the QPC. The backscattered wave arriving at the QPC will interfere irregardless of the finite temperature.

However, even though the interference pattern will not be destroyed, the magnitude of the backscattered wave will decay as $1/r^2$, where r is the radial distance away from the QPC. This is because only $1/r$ of the initial wave packet will be impinging on the impurities in general and the impurities pretty much acts as a s-wave scatterer, so it scatters uniformly in all direction, causing an additional decay factor of $1/r$.

2.5 Quantum Explanation of the Observed Stability

In this section, we will provide a quantum interoperation of the observed stability based on the thermal wavepacket approach and we will also introduce a classical interpretation at a later section.

Going back to the Gedanken experiment mentioned before, consider a pair of side-by-side QPCs differing by a shift. Suppose we put wavepacket A though one QPC and wavepacket B through the other. Can the coherent overlap between the initially nonoverlapping A and B wavepackets increase over time and distance from the QPC's? The answer is of course no, both classically (considered as overlap in phase space) and quantally. It is elementary to show that the coherent overlap must remain the same over time if the wavepackets are propagated under the same Hamiltonian. This is true whether or not disorder is present.

However, in the experiment as performed, the Hamiltonian of a single QPC and the Hamiltonian with the QPC shifted over are not the same. Therefore, no theorem constrains the evolution of the coherent overlap between the two different initial wavepackets. As it will become clear, this is exactly what leads to the observed stability.

In order to show this, we first consider the ideal case where the QPC is perfectly adiabatic and provide an analytical solution of the coherent overlap between the two wavepackets launched from two different QPCs as a function of time. We show the correlation reaches almost one at sufficiently large distance even if the initial coherent overlap is negligible. We then choose a more realistic QPC potential and also add smooth disorder of the type causing the branching into the system. The coherent overlap in this case still reaches 85%. Finally, we show that the same mechanism works for classical trajectories. For the classical case, we calculate an overlap of 79% in the phase space. Both results prove that the stability in the experiment is due to the nature of the experimental QPC shift. In the last part of the chapter, we also make a prediction on the stability of the branched flow when the second mode in the QPC is open.

For a QPC with harmonic confinement, the Hamiltonian is

$$H_o = \frac{\vec{p}^2}{2m} + \frac{1}{2}m\omega^2(y)x^2, \tag{2.10}$$

where $\omega(y)$ is a slowly varying function of y and decreases monotonically as the QPC opens up. According to the thermal wavepacket approach [10], we can reproduce the experimental results by propagating an initial wavepacket of the following form through the system.

$$\Psi_o(x,y,0) = \int dE\, exp(i\phi(E,y_0))\sqrt{-\frac{m}{2\pi\hbar^2}\frac{\partial f_T(E,E_F)}{\partial E}}\Psi_1(x,y,E) \quad , \tag{2.11}$$

where $\Psi_1(x,y,E)$ is the scattering eigenstate at energy E , $\phi(E,y_0)$ is chosen so that it is a compact wavepacket centered at $y = y_0$, and $f_T(E,E_F)$ is the Fermi distribution with temperature T and Fermi energy E_F. This is the thermal wavepacket at temperature T. For one mode open in the QPC, the thermal wavepacket gives the correct thermally averaged conductance, by propagating the wavepacket through the scattering region and counting the total flux that passes through a given point. Because of the low energy uncertainty at low temperature, the thermal wavepacket is usually very broad. In order to make the numerical simulation more efficient, we create a waveguide with one open mode in front of the QPC and channel it through the center of the QPC. y_0 is chosen to be inside the waveguide such that majority of the thermal wavepacket lies inside the waveguide before the QPC.

First we consider a perfect QPC and assume that disorder is absent. Numerical results including both disorder and an imperfect QPC follow. For a perfectly adiabatic QPC satisfying $\frac{\omega'(y)}{\omega^2(y)} \ll \frac{m}{\hbar k_F}$ and $\frac{\omega''(y)}{\omega^2(y)} \ll \frac{m}{\hbar}$, where k_F is the Fermi wavevector, $\Psi_1(x,y,E)$ can be approximated by

$$\Psi_1(x,y,E) = \frac{A(y_0)}{\sqrt{\hbar k(y,E)}\sqrt{\pi}\sigma(y)}exp(i\int_{y_0}^{y} k(y',E)dy' - \frac{x^2}{2\sigma(y)^2}), \tag{2.12}$$

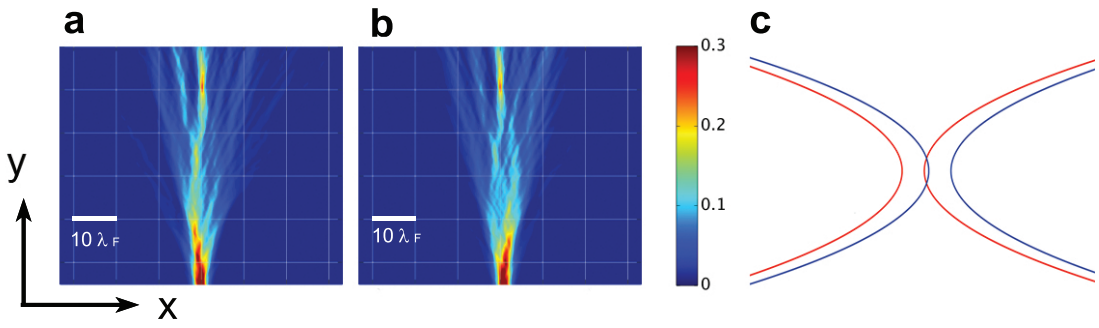

Figure 2.2: Quantum simulations of the total flux in the y direction that passes through a given point. (**a**) plots the case where the QPC is not shifted and (**b**) shows the case where the QPC is shifted by λ_F to the right . The white reference grid($10\lambda_F$) denotes the same location in both images. The color axis shows the normalized density of flux per wavelength. Both images start at y=$y_0 + 15\lambda_F$. (**c**) A schematic plot of the QPCs used in the simulations(Red in (**a**) and blue in (**b**)).

where $\sigma(y) = \sqrt{\frac{\hbar}{m\omega(y)}}$, $\frac{\hbar^2 k^2(y,E)}{2m} + \frac{1}{2}\hbar w(y) = E$ and $A(y_0)$ is the normalizing constant.

The effect of shifting the QPC is incorporated in the initial wavepacket as

$$\Psi_s(x,y,0) = \hat{L}(x_0)\Psi_o(x,y,0), \tag{2.13}$$

where $\hat{L}(x_0) = e^{-ix_0\hat{p}_x/\hbar}$ is the translation operator, x_0 is the displacement of the QPC and $\hat{p}_x$ is the momentum operator in the x direction.

The two wavepackets evolve under the influence of their respective QPC and the coherent overlap between them at a later time t is given by

$$\begin{aligned} C_{o,s}(t) &= \left| \int dxdy\ \Psi_o^*(x,y,t)\Psi_s(x,y,t) \right| \\ &= \left| \int dy\ H(y,t)S(y) \right| \end{aligned}, \tag{2.14}$$

and

$$H(y,t) = \int dEdE' \, \frac{|A(y_0)|^2 \, a^*(E')a(E)}{\hbar\sqrt{k(y,E)k(y,E')}} exp(-i(E-E')t/\hbar) \times e^{i(\phi(E,y_0)-\phi(E',y_0)+\int_{y_0}^{y}(k(y',E)-k(y',E'))dy')} \quad . \tag{2.15}$$

$$S(y) = exp(-\frac{x_0^2}{4\sigma(y)^2})$$

H(y,t) is essentially a function needed for normalization and the integral can be estimated by considering only S(y). When x_0 is 0, $S(y) = 1$ and normalization guarantees that $C_{o,s}(t) = 1$. Initially, the wavepackets are centered around $y = y_0$ and we could choose an initial displacement $x_0 \gg \sigma(y_0)$ such that $S(y_0) \sim 0$ and $C_{o,s}(0) \sim 0$. As time increases, the wavepacket will move away from the injection point and broaden. At typical experimental temperatures, the broadening is small compared with the distance it travels in y[10]. When the centers of the wavepackets reach a region far from the the injection point, $\sigma(y)$ around the new centers will grow to be much larger than x_0 and we have $S(y) \sim 1$, $C_{o,s}(t) \sim 1$. In other words, even though we start with two almost nonoverlapping(incoherent) initial wavepackets, the QPCs increase the coherent overlap as the wavepackets move away and this coherent overlap can reach unity in the far region. This overlap is of coherent nature and is different from the trivial spatial overlap one might expect. Spatial overlap is not enough to explain the experimentally observed stability due to the fluctuating phase in chaotic systems. However, our result shows that the overlap is large even if one takes into account the phases of the wavepackets and this coherent overlap can not be destroyed by disorders. This large coherent overlap only exists because the two different QPCs represent two different Hamiltonians. If propagated under the same Hamiltonian, the coherent overlap will always remain small.

In the experiment, both the finite size of the QPCs (making them not perfectly adiabatic) and the disorder can degrade the coherent overlap. We numerically estimate the coherent overlap under these conditions. The QPC's size is estimated from the Scanning Gate Microscopy

data in[9] and the random potential has a correlation length of $0.9\lambda_F$ and standard deviation of $8\%E_F$, where λ_F is the Fermi wavelength and $E_F = 7.5meV$ is chosen to match that in the experiment. The random potential is generated to match both the sample mobility and the distance from donors to 2DEGs[9]. The numerical results show that $C_{o,s}(t)$= 85% when the QPC is shifted by λ_F as in the experiment. Thus the degradation of the coherent overlap at long range from the QPC is modest.

Starting with two almost nonoverlapping(incoherent) initial wavepackets, and evolving separately under the influence of two different QPCs, their coherent overlap increases with time and distance, increasing fastest close to the QPCs. The coherent overlap eventually saturates to some constant value far from the QPCs. The two wavepackets now evolve effectively under the same Hamiltonian and their coherent overlap cannot be changed by the presence of disorder, for example. This is why we measure $25\lambda_F$ downstream from the injection point, where the potential due to the QPC has died off. *Given the large coherent overlap between the two wavepackets, we should expect the same set of branches far from the injection point even though the flow patterns look different close to the QPC, as shown in our quantum simulations in Fig.2.2*.

Another kind of disorder can reduce the overlap: backscattering from hard impurity scatterers. However, backscattering was suppressed in the original experiment due to the high purity of the samples used[9].

One can also formulate this problem in the language of the traditional scattering theory in the follwing way:

$$\begin{aligned} \Psi_o(x,y,t) &= \int dE\, a(E) e^{-iEt/\hbar} \sum_l T_{o,l}(E)\Psi_l(x,y,E) \\ \Psi_s(x,y,t) &= \int dE\, a(E) e^{-iEt/\hbar} \sum_l T_{s,l}(E)\Psi_l(x,y,E) \end{aligned}, \tag{2.16}$$

where $\Psi_l(x,y,E)$ is one of the scattering eigenstates labeled by l in the asymptotic region with energy E and $T_{o,l}(E)$ and $T_{s,l}(E)$ are the transmission coefficients. The overlap is given by

$$C_{o,s}(t) = \left| \int dE \ |a(E)|^2 \sum_l T^*_{s,l}(E) T_{o,l}(E) \right|, \tag{2.17}$$

and it is mainly generated by the QPCs rather than the random potential.

2.6 Classical Explanation

Reference[9] included both classical and quantum simulations and discussion. Does our explanation of the branch stability also apply to classical simulations? Indeed it does, but the proper classical initial conditions to represent the QPC are subtle and require care. A choice that closely resembles the quantum initial conditions is to use the Wigner quasiprobability distribution[13], defined as

$$P(\vec{x},\vec{p}) = \frac{1}{\pi\hbar} \int_{-\infty}^{\infty} d\vec{s}\, e^{2i\vec{p}\cdot\vec{s}/\hbar} \Psi^*(\vec{x}+\vec{s})\Psi(\vec{x}-\vec{s}). \tag{2.18}$$

The advantages are twofold: a) it produces the correct quantum spatial distribution $P(\vec{x}) = \int P(\vec{x},\vec{p})d\vec{p} = |\Psi(\vec{x})|^2$ and momentum distribution $P(\vec{p}) = \int P(\vec{x},\vec{p})d\vec{x} = |\Psi(\vec{p})|^2$; b) it properly accounts for the momentum uncertainty due to the confinement of QPC. Keeping y fixed at y_0, applying (2.18) to (2.12) in x yields

$$P(x,p_x) = \frac{1}{\pi\sigma_{p_x}\sigma_x} exp(-\frac{p_x^2}{\sigma_{p_x}^2} - \frac{x^2}{\sigma_x^2}), \tag{2.19}$$

where $\sigma_{p_x}^2 = m\hbar\omega(y_0)$ and $\sigma_x^2 = \hbar/m\omega(y_0)$.

When $\omega(y)$ changes sufficiently slowly compared to the motion in y, (2.19) holds approximately true for any $y > y_0$(where the y_0 should be replaced by y in the equation), which

implies that momentum distributions are highly correlated at any position no matter which QPC the electron originates from. The only difference is the overall probability of arriving at that point. This already hints as to why branches remain at the same positions with a modified strength.

As in the quantum case, we use a realistic QPC potential, and weak random potentials in the open regions of the 2DEGs. We sample according to (2.19), with the keeping energy fixed at E_F by eliminating trajectories with larger energy in the Wigner distribution, and boosting those with less in p_y. These details may be omitted and do not change the conclusions about branch populations and overlap. We propagate the electrons classically. We show Poincare' surface of section plots[11,17] in Fig.2.3. In the absence of disorder(Fig.2.3a and b), the adiabaticity of the QPC ensures that when the electrons emerge, most energy is transferred from the x (transverse) direction to the y (longitudinal) direction, which is also expected in the quantum case. The results when disorder is present is shown in Fig.2.3c&d. As can be seen, very similar regions in phase space are occupied, with different relative strengths, when the QPC is shifted. To quantify the overlap in phase space, we define the correlation to be

$$C(P_o, P_s) = \int dx dp_x \sqrt{P_o(x, p_x) P_s(x, p_x)}, \tag{2.20}$$

where P_o corresponds to the distribution in the original QPC and P_s the shifted one. Twenty five wavelengths away from the injection point, it is measured that $C(P_o, P_s)$=79%, which is comparable to our quantum result.

The classical approach to branching is based on caustics which develop in coordinate space due to focussing effects, and stable regions in phase space[14,15,16] which persist some distance away from the injection point. (Eventually, stable regions, which form by chance so to speak in the random potential, are also subject to destruction further on in the random potential). Each branch corresponds to a localized region in phase space with its strength determined by

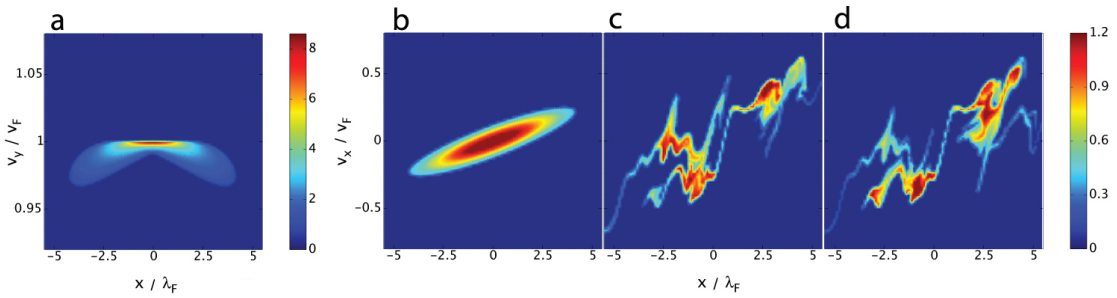

Figure 2.3: Poincare Surface of Section at y=y_0+25λ_F. (**a**) & (**b**) are calculated in the absence of disorder, but disorder is present in (**c**) & (**d**). (**c**) is plotted before the QPC is shifted and (**d**) is when the QPC is shifted by λ_F to the right. The color axis shows $\sqrt{P(x,v_y)}$ in (**a**) and $\sqrt{P(x,v_x)}$ in the rest, all in unit of $1/\sqrt{v_F\lambda_F}$, where v_F is the Fermi velocity and v_x and v_y are the velocity in the x and y direction respectively.

the electron density in those regions. After shifting the QPC, similar regions in phase space are occupied with only a changed relative density, which means in coordinate space that the same branches are occupied with a different strength. This explains the observation in the experiment[9].

In Fig.2.4, we present our simulations of the total classical flux that passes through a given point, which confirms that classical trajectories can indeed reproduce the observed stability. It is worth noting that the same stable regions could in principle be populated from both QPCs, causing some similarity of branch appearance, but this will not be a generic effect for all random potentials and QPC shifts. If we have a stable region in phase space that happens to be cut into halves by the shifting of the QPC, it will produce close branches with similar strength even with zero overlap at all times. However, there are two reasons why this will only have marginal effect in producing the experimentally observed stability. First, the probability of cutting a stable regions in halves such that both regions have appreciable strength is negligible. Therefore, it is not reproducible by experiment. Secondly and most importantly, any classical stable regions will decay with distance. Thus, any close branches, if exist, will separate eventually and contribute only marginally to the experimental stability.

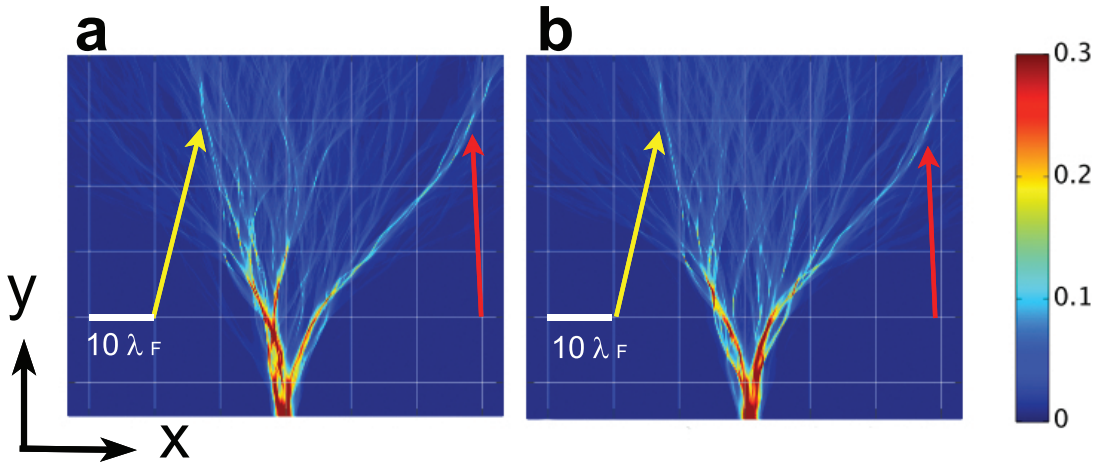

Figure 2.4: Classical simulations of the total flux in the y direction that passes through a given point. (**a**) shows the case where the QPC is not shifted and (**b**) shows that when the QPC is shifted by λ_F to the right. The white reference grid is at the same location in both images and the plots start at y=$y_0 + 15\lambda_F$. In both figures, the branches labeled by the red and yellow arrows are clearly visible.

2.7 Additional Stability with regards to different modes of QPC

One advantage of a classical interpretation is that it can provide intuition in cases where the quantum dynamics is less intuitive. One example would be to consider what happens when the second mode of QPC is open. According to reference [10], we need to independently propagate two wavepackets where one corresponds to the first mode and the other corresponds to the second mode. Their contributions to the flux are then added up incoherently to produce the experimental measurements. Since the contribution from the first mode is added incoherently, it is no surprise that the same set of branches recurs when both modes are open. However, it is interesting to ask what happens if one looks at the contribution from each mode alone. Quantum mechanically, the first and second mode are orthogonal to each other, and, therefore have zero overlap at all time since the Hamiltonian is the same. However, the classical phase space regions corresponding to the second mode alone would still overlap more or less with that due to the first mode. This doesn't depend on the choice of the Wigner quasiprobability distribu-

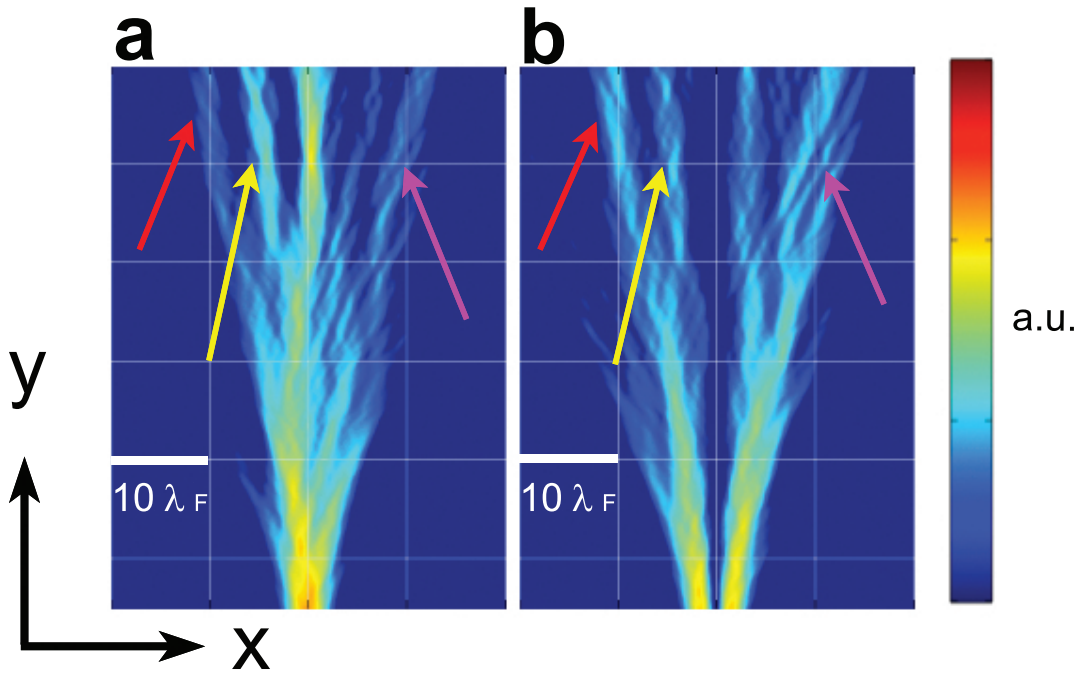

Figure 2.5: Quantum simulations of total flux in the y direction that passes through a given point when the second mode of the QPC is open. (**a**) corresponds to the first mode of the QPC while (**b**) shows contribution from the second mode alone. The starting point and length scale are the same as in Fig.2.2, but the flux strength is presented in log scale instead. In both figures, the branches pointed to by the red, yellow and purple arrows are clearly visible.

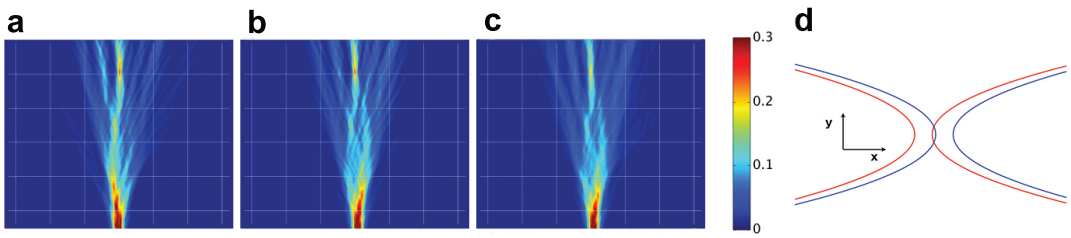

Figure 2.6: Quantum simulations at different temperature. The temperature is 0K in **a** & **b**, and 20K in **c**. **a** is plotted when the QPC is not shifted. **b** and **c** show the pattern when the QPC is shifted by λ_F to the right. The grid spacing is $10\lambda_F$ and the color axis shows the normalized denisty of flux per wavelength. All plots start at y=$y_0 + 15\lambda_F$.(**d**) A schematic plot of the QPCs used in the simulation(Red in **a** and blue in **b** and **c**.

tion. It is simply due to the fact that both modes have to live in the same region in coordinate space and maintain considerable amount of momentum uncertainty due to the confinement. As a result, the second mode alone should still produce some similar branches that appear in the first mode with a different strength. In order to see this effect, we take the logarithm of the quantum flux due to each mode alone and present the results in Fig.2.5. As we can see, some of the strongest branches are clearly preserved, which verifies our prediction.

2.8 Additional Stability: Energy Changes

For an individual classical electron in a chaotic system, its classical trajectory is exponentially unstable to perturbation. In the previous sections, the perturbation comes from the changes in the initial positions of the electrons. A different type of perturbation comes from the uncertainty in the electron's energy.

For experimental setup, the electron's energy is determined from the electron density, which depends on the density of charged atoms. One way to add energy uncertainty into the system is by changing the temperature of the system.

In the experiment, the random potential in the simulation is generated using screened donor potential with standard deviation $8\%E_F$ and correlation length $l_c = 0.9\lambda_F$[9,11]. The original

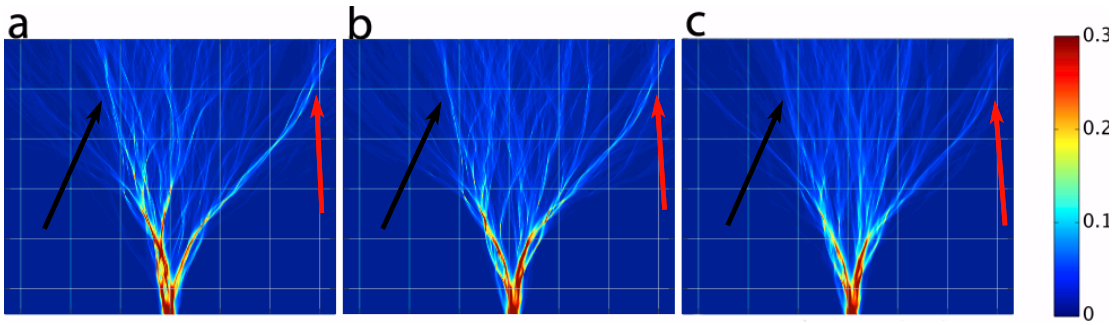

Figure 2.7: Classical simulations at different energies. Energy is fixed at E_F in **a** & **b**. In **c**, the energy distribution is chosen to match that of a thermal wavepacket at T=20K. **a** is plotted when the QPC is not shifted. **b** and **c** show the pattern when the QPC is shifted by λ_F to the right. In all three figures, the branches labeled by red and black arrows are clearly visible. The grid spacing is $10\lambda_F$ and the color axis shows the normalized denisty of flux per wavelength. All plots start at y=$y_0 + 15\lambda_F$.

experiment was performed at 4.2K[9] with thermal length $2.8\lambda_F$[10]. At 20K, the thermal length is reduced to $0.6\lambda_F < l_c$. For different scattering amplitude to interfere effectively, the scatterers need to be within half thermal length[10]. If quantum interfence from multiple scattering off disorders is involved in producing the stable flow pattern in experiment as originally suspected[9], we would expect thermal averaging to change the flow pattern significantly and smooth away the stability at 20K. However, except for strength changes, finite temperature does little damage to the location and shape of branches in Fig.2.2c, which also holds true for classical trajectories in Fig.2.4c.

This additional stability to thermal averaging is different from the experimentally observed stability and can't be explained by the overlap idea. In this case, the Hamiltonian is the same and the only difference is temperature or equivalently the intial thermal wavepacket. For two initial thermal wavepackets with $T = T_1$ and $T = T_2$, their overlap is not changing over time since they evolve under the same Hamiltonian. Holding T_1 fixed and greater than 0, the overlap can be made arbitrarily small as $T_2 \to 0$. Thus, this additional stability is not produced by overlap and different from the experimentally observed one.

More surprisingly, this stability is even shared by classical branched flow as well. In this

case, the electrons start at different places in the classical phase space and they remain separated in the phase space through the whole process. This is because now the Hamiltonian is the same and electrons with different energy remain separated in the phase space and their path in the phase space will never cross.

Classical theories can be used to explain this additional stability. As a result of energy quantization in the transverse direction, energy uncertainy due to finite temperature is completely borne by the logitudinal direction. However, caustics and stable regions are mostly related to the motion in the transverse direction[16] and changes in v_y has little effect on the structure in phase space. In Fig.2.8, we start with trajectories that have the same distribution in v_x, but with 22% energy variation($kT/E_F = 22\%$ at 20K), and plot the surface of section at $y = y_0 + 5l_c$. As expected, changes in v_y only slightly shift the surfaces and does little damage to where the density builds up.

2.9 Summary

Branching is a universal phenomenon of wave propagation in a weakly correlated random medium. It is observed in 2DEGs with wavelengths on the scale of nanometers[1,2], in quasi-two-dimensional resonator with microwave[4] and used to study sound propagation in oceans with megameter length scales[5]. It has significant influence on electron transport in 2DEGs[6,7] and is found to be implicated in the formation of freak waves in oceans[8]. In all these studies, classical trajectory simulations show closely similar branch formation.

However, the classical interpretation was challenged by a recent experiment on 2DEGs[9], where it was observed that the (necessarily quantum) branched flow pattern even far away showed stability of the branches against the changes in the QPC. This stability was conjectured to be of quantum origin[9]. To our best knowledge, no insights into this stability have been provided since, and it remains a puzzle in the literature. In this chapter, we provide an

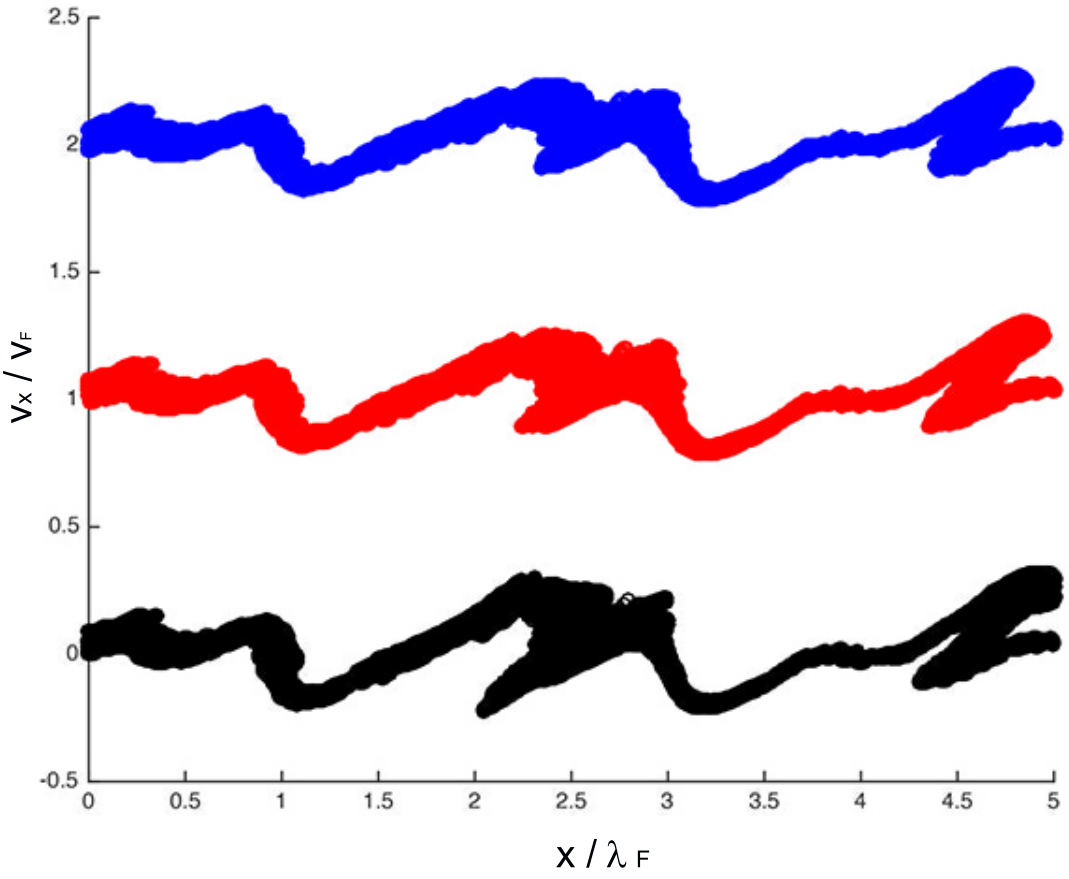

Figure 2.8: Poincare Surface of Section for different energies. Poincare Surface of section at $y = y_0 + 5l_c$. The respective energy is $0.78E_F$(Blue), E_F(Red) and $1.22E_F$(Black), and all initial manifolds start with v_x uniformally distributed between $[-0.05v_F, 0.05v_F]$ and x ranging from $-10\lambda_F$ to $10\lambda_F$. In this simulations, the QPC is absent and the electrons are initially placed in free space before the random potential($y = y_0$). In the figure, we only take a snapshot of the phase space structure with x between $[0, 5\lambda_F]$ and the curves are shifted from each other vertically by one unit.

explanation for the observed stability. Moreover, we provide numerical simulations to show that it can indeed be reproduced by using only classical trajectories.

In conclusion, we have successfully explained the stability of branched flow against large changes in initial conditions using both quantum and classical simulations, which agree on the fact of the stability of branches against shifts of the QPC injection point. This resolves a puzzle raised by a recent experiment[9] and shows the role of the QPC in enhancing the stability of branched flow in 2DEGs. Our classical interpretation predicts a further stability of the branched flow of the second mode in the QPC that can not be readily inferred from the experiment. The interpretations in this chapter can provide useful insights into future applications in the coherent control of electron flow, branch management and probing local random potential.

3

Light Scattering by Metallic Nanoparticles and Three Types of Plasmon: Background

In this chapter, we will switch gear to talk about the second main topic of this thesis: Localized Surface Plasmon.

In a nutshell, the study of Localized Surface Plasmon is about light scattering by metallic particles whose dimensions are at nanoscales. What makes it interesting is that the scattering here is due to the resonant coupling between the incoming light and the collective excitation of the conduction electrons in the metallic nanostructures.

In this chapter, we give a brief review of the physics background of plasmons and for more detailed information, we refer the readers to the book written by Prof. Stefan Maier[19].

3.1 Brief Review of Electromagnetism

The theory of plasmons can be divided into two parts: a microscopic theory explaining the electromagnetic response of the electrons in the metallic nanostructure, which gives rise to a macroscopic dielectric constant that characterizes the nanostructure, and a macroscopic theory describing the electromagnetic response of the nanostructure to an incident light. This macroscopic theory is based on the classical Maxwell's equations.

3.1.1 Maxwell's equations

The Maxwell's equations are among the most fundamental and elegant equations of physics and they are the foundation of classical electromagnetism. In its differential form, they are given by[44]

$$\begin{aligned} \bigtriangledown \cdot \vec{D} &= \rho_{ext} \\ \bigtriangledown \cdot \vec{B} &= 0 \\ \bigtriangledown \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\ \bigtriangledown \times \vec{H} &= \frac{\partial \vec{D}}{\partial t} + \vec{J}_{ext} \end{aligned} \quad . \tag{3.1}$$

$\vec{E}$ is the electric filed, $\vec{D}$ is the dielectric displacement, $\vec{B}$ is the magnetic induction and $\vec{H}$ is the magnetic field. ρ_{ext} is the external or free charge density and $\vec{J}_{ext}$ is the external current density. These two terms can be thought as the source of the electromagnetic fields.

For a homogenous material with linear response, we can express $\vec{H}$ and $\vec{D}$ as

$$\begin{aligned} \vec{D} &= \epsilon_0 \epsilon \vec{E} \\ \vec{B} &= \mu_0 \mu \vec{H} \end{aligned} . \tag{3.2}$$

Given the above six equation, together with the appropriate boundary conditions, one can in theory solve any problems in electromagnetism. In its essence, Maxwell's equations are a phenomenonical approach to solving scattering problems in electromagnetism where the microscopic response of the materials is summarized by ϵ and μ. In order for one to solve the problem, one needs to be given both ϵ and μ. This could be either obtained from an experimental measurement or a theoretical calculation taking into account the atomic structure of the material.

In general, a material's electromagnetic response will depend on both the frequency and the wave vector of the incident wave, so Eq.(3.2) is better rewritten as

$$\begin{aligned} \vec{D}(\vec{k},\omega) &= \epsilon_0 \epsilon(\vec{k},\omega)\vec{E}(\vec{k},\omega) \\ \vec{B}(\vec{k},\omega) &= \mu_0 \mu(\vec{k},\omega)\vec{H}(\vec{k},\omega) \end{aligned} , \tag{3.3}$$

where $\vec{E}(\vec{k},\omega)$ is the amplitude of the normal mode with wave vector $\vec{k}$ and frequency ω.

Since metallic nanoparticles are non-magnetic materials, we shall focus only on the electric part of the equations. In other words, we shall assume that $\mu(\vec{k},\omega) = 1$ for the case we are interested in.

For metal, we can further write the current density as a linear response of the electron motion to the total electric field by introducing a coefficient called conductivity, that is

$$\vec{J}(\vec{k},\omega) = \sigma(\vec{k},\omega)\vec{E}, \tag{3.4}$$

where $\sigma(\vec{k},\omega)$ is the dynamic conductivity of the material.

There is a deep connection between the dynamic conductivity of a material and its dynamic polarizability. This should make intuitive sense since both are directly related to the motion of the conduction electrons.

To see this connection, we can express the current density as a function of polarization density:

$$\vec{J}(\vec{k},\omega) = \frac{\partial \vec{P}}{\partial t}, \tag{3.5}$$

where $\vec{P}$ is the polarization density. This relation is most easily understood when one considers a single electron with position $\vec{x}$, such that $\vec{P} = -e\vec{x}$. Thus, $\frac{d\vec{P}}{dt} = -e\vec{v}$. Multiplying both sides by the electron density will give the relation in (3.5).

In terms of the dielectric constant, $\vec{P}$ can be related to $\vec{E}$ as

$$\begin{aligned} \vec{P} &= \vec{D} - \epsilon_0 \vec{E} \\ &= (\epsilon - 1)\epsilon_0 \vec{E} \end{aligned}. \tag{3.6}$$

Furthermore, for normal modes, we can replace $\frac{\partial}{\partial t}$ by $-i\omega$. Combining (3.4), (3.5) and (3.6), we get a relation between the conductivity and polarizability

$$\begin{aligned} \sigma(\vec{k},\omega)\vec{E}(\vec{k},\omega) &= -i\omega\vec{P} \\ &= -i\omega(\epsilon - 1)\epsilon_0 \vec{E}(\vec{k},\omega) \end{aligned}. \tag{3.7}$$

Thus,

$$\epsilon(\vec{k},\omega) = 1 + \frac{i\sigma(\vec{k},\omega)}{\epsilon_0 \omega}. \tag{3.8}$$

As one probably remembers from his introductory electromagnetism class, the real part of the dielectric constant decides the dispersion relation in the case of small dissipation, while the imaginary part of the dielectric constant determines the dissipation. This makes more intuitive

sense when looking at (3.8). In (3.8), the imaginary part of the dielectric constant is proportional to the real part of the conductivity.

When the conductivity is purely real, it means that the current is always in phase with the driving electric field. Since current is nothing more than electrons moving with a velocity, this translates into the fact that electrons moves in phase with the electric field such that the electric field always acting as an energy source trying to accelerate electrons. This energy input has to be balanced by an energy output due to relaxation process such as electron phonon scattering. If one uses a relaxation time to characterize this relaxation process, it leads to the simple yet usually sufficient Drude model for conductivity. Thus, it is not hard to see why the real part of the conductivity will lead to dissipation. This loss is essentially resistive loss as one encounters in everyday life: the light bulb heats up as one sends electricity through it.

On the other hand, when the dielectric constant is purely real, we can see from (3.5) that the current and the electric field will differ by a phase of $\pi/2$. In one optical cycle, the electrons will be in phase with the electric filed and out of phase with the electric filed in the other half of the cycle. When they are in phase, the electrons pull energy from the electric field and when they are out of phase, the electrons convert its gained kinetic energy into the potential energy. On average, there is no energy exchange between the electric field and the electrons in this perfect dielectric material. However, the light will still interact with the electrons, leading to changes in the velocity of the light propagating through the sample, i.e. a different dispersion relationship.

For most materials, the dielectric constant/conductivity will not be purely real and it is in general a complex number involving both dielectric type of behavior and conductor type of behavior.

Also, for the frequency range we are interested in, the wavelength of the incident light is usually much larger than atomic spacing. In this long wavelength regime, we can assume that

the material is isotropic and homogenous, and take the limit that $\vec{k} = 0$. As a result, we can rewrite (3.8) as

$$\epsilon(\omega) = 1 + \frac{i\sigma(\omega)}{\epsilon_0 \omega}. \tag{3.9}$$

In the following sections, we will only discuss in terms of $\epsilon(\omega)$.

3.1.2 Drude model for conductivity

One popular and powerful model for conductivity is the Drude model. It is specifically designed to characterize the response of free electron gas to an external driving electric field.

Assuming weak electron-electron interaction, the equation of motion for a single electron in an external electric field is given by

$$m\ddot{\vec{x}} + m\frac{\dot{\vec{x}}}{\tau} = -e\vec{E}, \tag{3.10}$$

where τ is the relaxation time.

If we focus on the electric response to normal modes, we can replace the time derivative with $-i\omega$ by assuming $\vec{x}(t) = \vec{x}_0 e^{-iwt}$. The solution to (3.10) then becomes:

$$\vec{x}(t) = \frac{e}{m(\omega^2 + i\omega/\tau)}\vec{E}(t). \tag{3.11}$$

By definition, if we have electron density n and each electron is displaced by the same amount $\vec{x}(t)$, the polarization density becomes

$$\vec{P}(t) = -ne\vec{x}(t) = \frac{-ne^2}{m(\omega^2 + i\omega/\tau)}\vec{E}(t). \tag{3.12}$$

Thus, the dielectric constant is given by

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega/\tau}, \tag{3.13}$$

where

$$\omega_p = \frac{ne^2}{m\epsilon_0}. \tag{3.14}$$

Surprisingly, this simple model works very well in practice. Part of the reason is that the relaxation time τ, which describes how fast the electron can lose energy it obtained from the electric filed, can usually be chosen to match experimental measurements.

3.2 Volume Plasmon

In this section, we introduce the first type of plasmons, the volume plasmon.

Volume plasmon is more popularly known as the ”plasmon” in the condensed matter community. It is the collective excitation or movement of electrons in an electron gas. Here, instead of going into the process of trying to quantizing the excitation, we will instead show how it arises naturally just from the Maxwell's equations.

According to Maxwell's equations, when there are no external currents, we find, by taking the curl of the first equation in the Maxwell's four equations, the following relation

$$\bigtriangledown \times \bigtriangledown \times \vec{E} = - \bigtriangledown \times \frac{\partial B}{\partial t} = -\mu_0 \frac{\partial^2 \vec{D}}{\partial t^2}. \tag{3.15}$$

Noting that

$$\bigtriangledown \times \bigtriangledown \times \vec{E} = \bigtriangledown(\bigtriangledown \cdot \vec{E}) - \bigtriangledown^2 \vec{E}, \tag{3.16}$$

Figure 3.1: A simplified model to explain the formation of Volume Plasmon. The red cloud corresponds to the fixed ion cloud and the blue cloud corresponds to the electron cloud. The electron cloud is displaced from the positively charged ion cloud by a distance of u.

the previous equation can be simplified to

$$\vec{k}(\vec{k}\cdot\vec{E}) - k^2\vec{E} = -\epsilon(\omega)\frac{\omega^2}{c^2}\vec{E}. \tag{3.17}$$

There are two possible solutions to this equation, a transverse mode and a longitudinal mode.

For the transverse mode, $\vec{k}\cdot\vec{E} = 0$ and (3.17) reduces to

$$k^2 = \epsilon(\omega)\frac{\omega^2}{c^2}, \tag{3.18}$$

which is the familiar dispersion relation we expect to see in a nonmagnetic effective medium.

There is also a possible longitudinal mode, where $\vec{k}\cdot\vec{E} = kE$. The only way to satisfy (3.17) in this case is to set

$$Re[\epsilon(\omega)] = 0. \tag{3.19}$$

The solution to this corresponds to the so-called volume plasmon and the solution to (3.19) is $\omega = \omega_p$.

One can understand the physical origin of this volume plasmon by looking at a potentially oversimplified model[19]. As shown in Fig.3.1, the material is charge neutral. Suppose now that all the electrons are displaced from the positively charged ion cloud by the same distance u, it will leave behind a positively charged background of ions. If we assume that the positively charged ions are immobile, then it will create a restoring force acting on the electrons trying to pull them back to the change neutral positions.

The restoring force can be easily calculated with the assumptions that 1) the sample is large so we do not need to worry about the boundary conditions; 2) all the electrons are displaced by the same distance. With these two assumptions, the electric field is uniformly and constant inside. Its value is

$$\vec{E} = \frac{ne}{\epsilon_0}\vec{u}. \tag{3.20}$$

This electric field uniformly acts on all the electrons and serves as the source of the resorting force. Each electron experiences the same restoring force and as a result, they move in a collective fashion. Their motion is driven by the following dynamic equation:

$$m\ddot{u} = -\frac{ne^2}{\epsilon_0}\vec{u}, \tag{3.21}$$

which is the familiar equation for a harmonic oscillator with frequency

$$\omega_p = \frac{ne^2}{m\epsilon_0}. \tag{3.22}$$

This is the simplest, maybe oversimplified model to explain the existence of volume plasmon. For more sophisticated models and the quantization of the plasmon, the reader could consult the book by Prof. Giuseppe Grosso and Prof. Giuseppe Pastori Parravicini[42].

Experimentally, the volume plasmon is usually probed using electron energy loss spectroscopy.

3.3 Surface Plasmon Polariton

In this section, we introduce the second type of plasmon, the Surface Plasmon Polariton(SPP).

Unlike volume plasmon, which is a bulk state, SPP is essentially a surface state that exists only at the interface between a metallic material and a dielectric material.

3.3.1 Dispersion Relation

In the absence of external charge sources and current sources, the Maxwell's equations reduce to

$$\begin{aligned} \bigtriangledown \cdot \vec{D} &= 0 \\ \bigtriangledown \cdot \vec{B} &= 0 \\ \bigtriangledown \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\ \bigtriangledown \times \vec{H} &= \frac{\partial \vec{D}}{\partial t} \end{aligned} \tag{3.23}$$

If we take the curl of the third equation, we get

$$\bigtriangledown \times \bigtriangledown \times \vec{E} = \bigtriangledown(\bigtriangledown \cdot \vec{E}) - \bigtriangledown^2 \vec{E} = -\mu_0 \frac{\partial^2 \vec{D}}{\partial t^2}. \tag{3.24}$$

Also the first equation can be written as

$$\begin{aligned} \bigtriangledown \cdot \vec{D} &= \bigtriangledown \cdot (\epsilon_0 \epsilon \vec{E}) \\ &= \epsilon_0 \epsilon \bigtriangledown \cdot \vec{E} + \epsilon_0 \bigtriangledown \epsilon \cdot \vec{E}, \\ &= \epsilon_0 \epsilon \bigtriangledown \cdot \vec{E} = 0 \end{aligned} \tag{3.25}$$

assuming a homogeneous material with linear response.

Thus, (3.25) reduces to

$$\bigtriangledown^2 \vec{E} + \epsilon k_0^2 \vec{E} = 0, \tag{3.26}$$

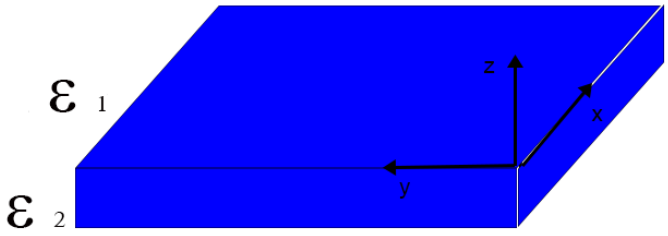

Figure 3.2: Dielectric Metal interface where SPP arises. It is the simplest geometry where SPP arises. The top half is a semi infinite dielectric material with dielectric constant ϵ_1 and the lower half is a semi infinite conductor with dielectric constant ϵ_2 .

where $k_0 = \frac{w}{c}$ is the wave vector of light in free space.

To derive the SPP modes, we need to make assumptions about the geometry of the system. For introductory purpose, we only consider a single layer system, where we assume that SPP arises in the interface between a semi-infinite dielectric material(usually air) and a semi-infinite conductor as shown in Fig.3.2. Multilayer systems are considered in Ref.[19].

We look for evanescent waves bound to the interface. By choosing the coordinate system, we can assume that the wave takes the following form

$$\vec{E} = \vec{E}(z)exp(i\beta x), \tag{3.27}$$

and (3.26) simplifies to

$$\bigtriangledown^2\vec{E}(z) + (\epsilon k_0^2 - \beta^2)\vec{E}(z) = 0. \tag{3.28}$$

Also, the third and fourth equations in (3.23) can also be written out explicitly in the follow-

ing form by assuming lack of dependence on y:

$$\begin{aligned} i\mu_0\omega H_x &= -\frac{\partial E_y}{\partial z} \\ -i\mu_0\omega H_y &= i\beta E_z - \frac{\partial E_x}{\partial z}, \\ i\mu_0\omega H_z &= i\beta E_y \end{aligned} \tag{3.29}$$

$$\begin{aligned} -i\epsilon_0\epsilon\omega E_x &= -\frac{\partial H_y}{\partial z} \\ i\epsilon_0\epsilon\omega E_y &= i\beta H_z - \frac{\partial H_x}{\partial z}. \\ -i\epsilon_0\epsilon\omega E_z &= i\beta H_y \end{aligned} \tag{3.30}$$

It can be shown[44] that once E_x and H_x are given, one can solve for the y and z components of the electromagnetic field.There are two sets of possible solution to the above six equations. One set is called the transverse magnetic(TM) mode, where $H_x = 0$, and the other set is called the transverse electric(TE) mode, where $E_x = 0$.

We first look for TM modes. Setting $H_x = 0$, we find that $E_y = 0$ and $H_z = 0$. The above six equations then reduce to

$$\begin{aligned} E_x &= -\frac{i}{\epsilon_0\epsilon\omega}\frac{\partial H_y}{\partial z} \\ E_z &= -\frac{\beta}{\epsilon_0\epsilon\omega}H_y \end{aligned} . \tag{3.31}$$

To determine H_y , we go back to (3.26) and it yields the following differential equation

$$\frac{\partial^2 H_y}{\partial z^2} + (\epsilon k_0^2 - \beta^2)H_y = 0. \tag{3.32}$$

Since we are looking for evanescent waves bound to the interface, we can assume that

$$H_y \propto exp(i\beta x - k_2 z), \tag{3.33}$$

for $z > 0$ and

$$H_y \propto exp(i\beta x + k_1 z), \tag{3.34}$$

for $z < 0$.

Taking advantage of (3.31), we get the following equations for $z > 0$:

$$\begin{aligned} H_y &= A_2 exp(i\beta x - k_2 z) \\ E_x &= \frac{ik_2}{\epsilon_0 \epsilon_2 \omega} A_2 exp(i\beta x - k_2 z) \quad . \\ E_z &= -\frac{\beta}{\epsilon_0 \epsilon_2 \omega} A_2 exp(i\beta x - k_2 z) \end{aligned} \tag{3.35}$$

Similarly, for $z < 0$, we have

$$\begin{aligned} H_y &= A_1 exp(i\beta x + k_1 z) \\ E_x &= -\frac{ik_1}{\epsilon_0 \epsilon_1 \omega} A_1 exp(i\beta x + k_1 z), \\ E_z &= -\frac{\beta}{\epsilon_0 \epsilon_1 \omega} A_1 exp(i\beta x + k_1 z) \end{aligned} \tag{3.36}$$

where k_1 and k_2 are the imaginary wave vector/decay constant in the z direction inside the conductor and the dielectric respectively.

To get the dispersion relation, we need to match the boundary conditions at the interface, which are given by the continuity of ϵE_z, E_x and H_y at the interface.

The condition $H_y(z = 0-) = H_y(z = 0+)$ requires that $A_1 = A_2$.

Similarly, $E_x(z = 0-) = E_x(z = 0+)$ requires

$$\frac{k_2}{k_1} = -\frac{\epsilon_2}{\epsilon_1}. \tag{3.37}$$

The two other equations needed to solve for k_1, k_2 and β are given by

$$\begin{aligned} k_1^2 &= \beta^2 - k_0^2\epsilon_1 \\ k_2^2 &= \beta^2 - k_0^2\epsilon_2. \end{aligned} \tag{3.38}$$

Combining these with (3.37), we find that

$$\beta = k_0\sqrt{\frac{\epsilon_1\epsilon_2}{\epsilon_2+\epsilon_1}}. \tag{3.39}$$

In the limit that $\epsilon_1 \to 0$, this gives $\beta = 0$, which is the long wavelength limit of the dispersion relation. According to (3.38), $k_1 = 0$ as well, meaning that the field can penetrate inside the metal without being decayed. Recalling that $\epsilon_1(\omega) = 0$ is exactly the condition required for the volume plasmon, this means that SPP becomes the volume Plasmon in this long wavelength limit.

The other set of solution is called the transverse electric mode(TE), where $E_x = 0$. Given $E_x = 0$, one can show that $E_z = 0$ and $H_y = 0$.

The above six equations then reduce to

$$\begin{aligned} H_x &= \frac{i}{\mu_0\omega}\frac{\partial E_y}{\partial z} \\ H_z &= \frac{\beta}{\mu_0\omega}E_y \end{aligned}. \tag{3.40}$$

The equation governing E_y is given by

$$\frac{\partial^2 E_y}{\partial z^2} + (\epsilon k_0^2 - \beta^2)E_y = 0. \tag{3.41}$$

Similarly to the TM modes, we look for surface states of the form

$$E_y \propto exp(i\beta x - k_2 z), \tag{3.42}$$

for $z > 0$ and

$$E_y \propto exp(i\beta x + k_1 z), \tag{3.43}$$

for $z < 0$.

For $z > 0$, the solution should take the following form because of the constraint set by (3.31)

$$\begin{aligned} E_y &= A_2 exp(i\beta x - k_2 z) \\ H_x &= -\frac{ik_2}{\epsilon_0 \epsilon_2 \omega} A_2 exp(i\beta x - k_2 z). \\ H_z &= \frac{\beta}{\epsilon_0 \epsilon_2 \omega} A_2 exp(i\beta x - k_2 z) \end{aligned} \tag{3.44}$$

Similarly, for $z < 0$, we have

$$\begin{aligned} E_y &= A_1 exp(i\beta x + k_1 z) \\ H_x &= \frac{ik_2}{\epsilon_0 \epsilon_1 \omega} A_1 exp(i\beta x + k_1 z), \\ H_z &= \frac{\beta}{\epsilon_0 \epsilon_1 \omega} A_1 exp(i\beta x + k_1 z) \end{aligned} \tag{3.45}$$

where k_1 and k_2 are the wave vector in the z direction inside the conductor and the dielectric respectively.

Similarly, the continuity condition for E_y at the interface requires that $A_1 = A_2$ and the continuity condition for H_x requires that

$$A_1(k_1 + k_2) = 0. \tag{3.46}$$

This condition could only be satisfied if $A_1 = 0$. Thus, for normal conductor, there is no TE SPP. However, for graphene, it is shown in Ref.[59] that one can find TE type of SPP modes. This is mainly due to the fact that graphene is a two dimensional metallic sheet that can provides external surface charges at the interface when graphene is sandwiched between two dielectric materials.

SPP modes can also be found in multilayer systems as well. The derivation is a little bit more complicated, but the spirit is the same.

3.3.2 Excitation Methods

One advantage of SPP is that it can compress the wavelength of the light. This can be inferred from (3.39). At high enough frequency, ϵ_1 can be made smaller than ϵ_2, leading to a compression in the wavelength of the SPP when compared to its wavelength in the dielectric material.

However, for silver and gold, there is a fundamental tradeoff between the field confinement and loss[19]. Usually, modes offering high confinement support only small propagation lengths, while modes offering high propagation lengths suffer from worse confinement. It is proposed[60] that the SPP modes in graphene can offer both large propagation lengths and large field confinement at the same time.

Due to the wavelength compression effect, there is an inherent wave vector mismatch between SPP and light in the dielectric material. Thus, SPP modes can not in general be excited by just shedding light onto the interface. There have been several methods proposed to resolve this wave vector mismatch[19].

The first method is called prism coupling[19]. This method takes advantage of the total internal reflection, where only evanescent waves get transmitted. Such evanescent waves can have high-k Fourier components, so it can excite the high-k SPP modes.

The secnd method is called grating coupling. The physics behind this method is very similar to Bragg scattering and one can excite the SPP modes resonantly when the following condition is satisfied

$$\beta = ksin\theta \pm ng, \tag{3.47}$$

where $g = 2\pi/a$ and a is the spacing between the grating and n is an integer.

The final method, which is probably the most popular method for both excitation and probing, is called Scattering-type Scanning Near-field Optical Microscopy(s-SNOM). It works by placing a metallic tip on top of the interface between the metallic material and the dielectric material. A focused laser beam is then guided to illuminate the metallic tip. This laser beam will excite a local electric dipole within the metallic tip and this electric dipole will then be able to excite the SPPs in the interface below through its near field. Compared with the previous method, s-SNOM enjoys the benefits of being able to excite both locally and also usually with high efficiency due to near field enhancement.

Another advantage of s-SNOM over the previous methods is its ability to be used as the probe as well. This is used to probe the SPP modes in graphene as done in Ref.[61] and Ref.[62]. The way this works is by probing the back scattered wave locally through exciting the metallic tip. In the experiment, an additional apparatus is usually employed to collect the backscattered wave from the tip. The collected backscattered wave is then brought to interfere with another reference beam to decide its amplitude and phase. As in the experiements[61,62], the graphene is cut into wedge shape and the edges can backscatter the SPP modes. When the backscattered SPP wave returns to the area below the tip, it could couple to the tip through near field coupling, which excite additional electric dipole field that can potentially reduce or enhance the original dipole field depending on its relative phase to the incoming light.

By moving the tip across the two dimensional interface and recording the phase and strength of the backscattered SPP wave, one can get information on the SPP modes. For instance,

in[61,62], the authors were able to determine the wavelength of SPP and its propagation length.

3.4 Localized Surface Plasmon

Finally, we will introduce the last kind of plasmons, Localized Surface Plasmon(LSP). LSP is the only type of plasmon we study in this thesis. It also arises from the interaction between light and the collective excitation of electrons in metal. The difference between LSP and SPP is that SPP is a surface state existing at the interface between a conductor and a dielectric material, while LSP is more like a resonant electric dipole. In some sense, SPP is more like a two dimensional surface state, while LSP is more like a zero-dimensional state.

LSP arises when one sheds light onto metallic nanostructures. As one of the simplest examples, we can consider light scattering by a single gold nanosphere. When the frequency of the incoming light coincides with the resonance frequency of the electrons moving inside the nanosphere, one can drive up a resonant scattering mode known as LSP.

Usually, for gold nanoparticles, the resonance wavelength of the light is on the order of 540nm, but the radius of the gold nanoparticles can be as small as 5nm. Since the wavelength of the light is two orders of magnitude larger than the size of the gold nanoparticles, one can employ what is known as the quasi-static approximation[19].

According to the quasistaitc approximation, one can solve for the scattering problem as an electrostatic problem, which ignores the phase fluctuation of the light across the gold nanosphere. The problem then becomes solving the Poisson equation with the appropriate boundary conditions set by the shape of the gold nanosphere

$$\bigtriangledown^2 \psi = 0, \tag{3.48}$$

where ψ is the static electric potential.

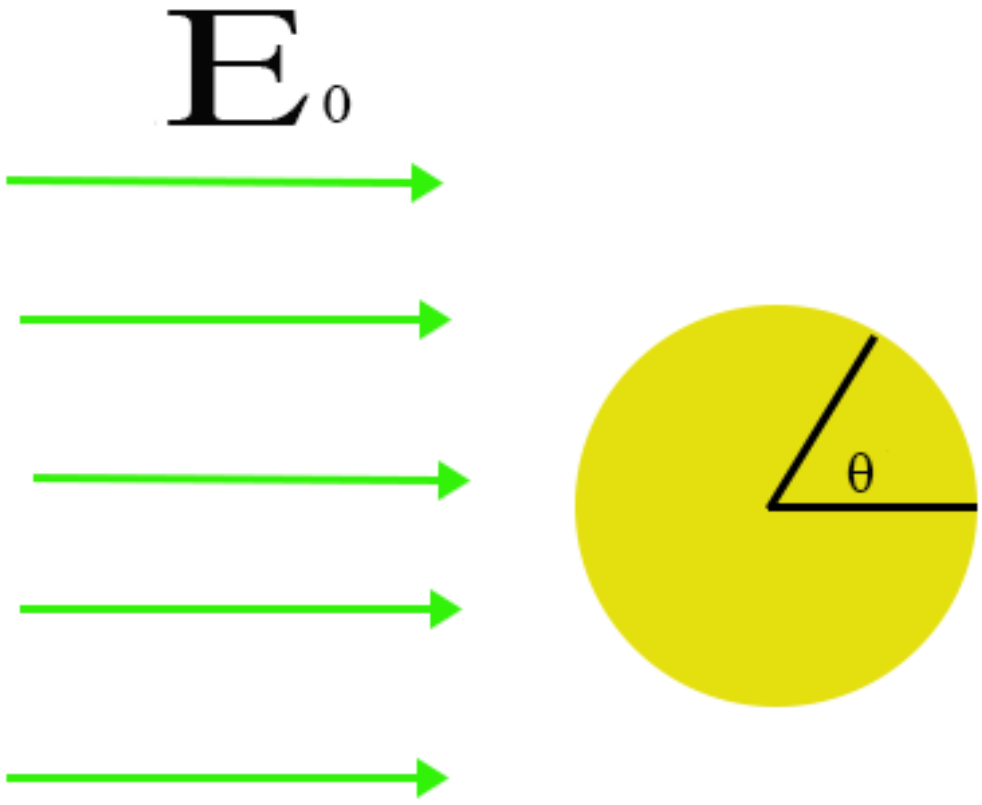

Figure 3.3: LSP within the quasi-static approximation. The green lines denote the static electric field and the yellow sphere is the gold nanosphere.

The incoming light can then be modeled as a uniform electric field. This problem can be solved analytically using the method of images[44]. The solution is given by

$$\begin{aligned} \psi_{in} &= -\frac{3\epsilon_m}{\epsilon + 2\epsilon_m} E_0 r cos\theta \\ \psi_{out} &= -E_0 r cos\theta + \frac{\epsilon - \epsilon_m}{\epsilon + 2\epsilon_m} E_0 a^3 \frac{cos\theta}{r^2}, \end{aligned} \tag{3.49}$$

where ψ_{in} is the electric potential within the nanosphere, ψ_{out} is the electric potential outside the nanosphere and a is the radius of the nanosphere. ϵ_m is the dielectric constant for the environment while ϵ is the dielectric constant of the metal, which can be a complex number. E_0 is the incoming electric field and θ is the angle defined in relation to the axis passing through the center of the nanosphere and parallel to the the direction of the electric field as shown in Fig.3.53.

(3.49) can be transformed into more physical form by noticing that the electric potential due to an electric dipole $\vec{p}$ is given by

$$\psi_p = \frac{\vec{p} \cdot \vec{r}}{4\pi\epsilon_0 r^3}. \tag{3.50}$$

Noticing the similarity between the second term of ψ_{in} and ψ_p, we can define an electric dipole moment as

$$\vec{p} = 4\pi\epsilon_0\epsilon_m a^3 \frac{\epsilon - \epsilon_m}{\epsilon + 2\epsilon_m} \vec{E}_0, \tag{3.51}$$

and we then can rewrite ψ_{out} as

$$\psi_{out} = -E_0 r cos\theta + \frac{\vec{p} \cdot \vec{r}}{4\pi\epsilon_0\epsilon_m r^3}. \tag{3.52}$$

This formula has a very simple physical interpretation. When we shed light onto the metallic nanosphere, the light induces an electric dipole moment within the metallic nanoparticles and the total electric field is essentially a summation of the contribution from the exciting field

and the scattering field due to the induced electric dipole moment.

Methods taking into account phase retardation and higher order modes can be employed to solve this scattering problem including Mie theory and Finite Difference Time Domain(FDTD) methods, but in general it is shown numerically that this simple quasistatic approximation works reasonably well for gold nanoparticles with characteristic length smaller than 100nm[41,43].

From (3.51), we can also see that one can get a resonance whenever the following condition is satisfied

$$Re[\epsilon] + 2\epsilon_m = 0. \tag{3.53}$$

In general, ϵ is a function of the frequency of the light, so we shall write it as $\epsilon(\omega)$. The frequency at which (3.53) is satisfied is called the LSP resonance frequency.

For the Drude model, $\epsilon_m(\omega)$ is given by

$$\epsilon(\omega) = \epsilon_\infty - \frac{\omega_p^2}{\omega^2 + i\omega/\tau}, \tag{3.54}$$

where ϵ_∞ is the correction for interband correction.

If the damping rate is small when compared with ω_p, the LSP resonance frequency after solving (3.53) is

$$\omega_{LSP} = \frac{\omega_p}{\sqrt{2 + \epsilon_\infty}}, \tag{3.55}$$

assuming that the dielectric environment is air.

Another quantity that is of relevance to experiments is the extinction cross section, which is a summation of the scattering cross section and the absorption cross section. They are given separately by

$$\begin{aligned} \sigma_{scattering} &= \frac{8\pi}{3} k^4 a^6 \left| \frac{\epsilon - \epsilon_m}{\epsilon + 2\epsilon_m} \right|^2 \\ \sigma_{absorption} &= 4\pi k a^3 Im[\frac{\epsilon - \epsilon_m}{\epsilon + 2\epsilon_m}] \end{aligned}. \tag{3.56}$$

One important point to notice here is that the scattering cross section scales as a^6, while the absorption cross section scales as a^3. Since the metallic nanoparticles are usually of small radius, it is therefore true that in most experimental setups, the extinction cross section is dictated by the absorption cross section.

4

Quantum Proximity Resonance and Linewidths of Localized Surface Plasmons

The previous chapter discusses the background for light scattering by metallic structures and introduces the three fundamental types of plasmons: volume plasmon, Surface Plasmon Polariton(SPP) and Localized Surface Plasmon(LSP).

In the rest of this thesis, we will focus only on LSP. We will draw an analogy between an interesting quantum multiple scattering phenomenon, known as the quantum proximity resonance, and plasmon dimer scattering.

4.1 Quantum Proximity Resonance

One of the key differences between quantum physics and classical physics in understanding electrons is that quantum mechanics interprets electrons as both wave and particles while classical mechanics interprets electrons as only particles.

For scattering theory, one consequence of the wave nature of electrons is that, besides the physical size of the scatterer, one has an additional metric for understanding the scattering strength of a scatterer and this metric is known as the scattering cross section. For classical scattering, what matters for a hard wall scatterer is its physical cross section. This makes intuitive sense. In the classical interpretation, an electron is modeled as a point mass, whose trajectory is a one dimensional topological line. For a hard wall potential, if the electron shoots in a direction that misses the physical cross section of the scattering ball, the scattering potential should not have an influence on this electron's trajectory.

However, the story is different according to quantum mechanics. It is known from most introductory quantum mechanics courses that the total scattering cross section for a hard wall sphere scattering is twice the physical cross section of the sphere.

The difference in the case of a hard wall sphere is due to the nonlocal wave nature of the electron scattering. Such difference can be made more significant when one adds resonance into the system. For a classical system with resonance in the sense of energy matching, the scattering cross section is still the same as its physical cross section. Imagine for instance a three dimensional harmonic oscillator, the classical scattering cross section is decided by the magnitude of the oscillation. If the point mass misses this physical cross section, it won't be scattered even in the presence of energy matching.

For quantum mechanics, electrons are interpreted as waves and as a wave, it also has a wavelength λ. Suppose that the scatterer(a three dimensional harmonic oscillator) has physical size of $a << \lambda$. The classical cross section is on the same order of magnitude as a^2 no

matter whether the resonance condition is satisfied or not. For quantum scattering without satisfying the resonance condition, the scattering cross section will be on the same order of magnitude as a^2, differing from the classical cross section by only a constant. However, if the resonance condition is satisfied, the scattering cross section in the quantum scattering case will scale as λ^2 [37].

Even if the wavelength is only one order of magnitude larger than the radius, the scattering cross section is two orders of magnitude larger. If one places two such "small" scatterers well outside each other's physical range, but within each other's scattering range, one can expect multiple scattering to take place. The phenomenon arising from this is known as the quantum proximity resonance [37], where if one places two resonant s-wave scatterers within one wavelength, it yields a sharp p-wave scattering mode with extremely small damping and a symmetric mode which is highly damped. Intuitively, for the p-wave scattering mode, the two s-wave scatterers are excited with exactly the opposite phase, so it almost will exactly cancel each other's far field, thus resulting in significantly reduced damping. However, for the symmetric mode, the two s-wave scatterers are excited in phase, so they enhance each other's far field through constructive interference, hence stronger damping.

4.2 Plasmon Dimer System

The collective excitation of conduction electrons in subwavelength structures is known as the Localized Surface Plasmon(LSP). Such plasmon modes have been intensively studied using noble metal nanoparticles [20,21,22,23]. More recently, the possibility of building terahertz metamaterials supporting such LSP modes is explored using graphene microribbons [24] and microdisks [25]. LSP holds promise for applications in ultrasensitive biosensing [27], nano-optical tweezers [29] and improved photovoltaic devices [30].

LSP arises when a subwavelength metal nanosphere is illuminated by light resonant with

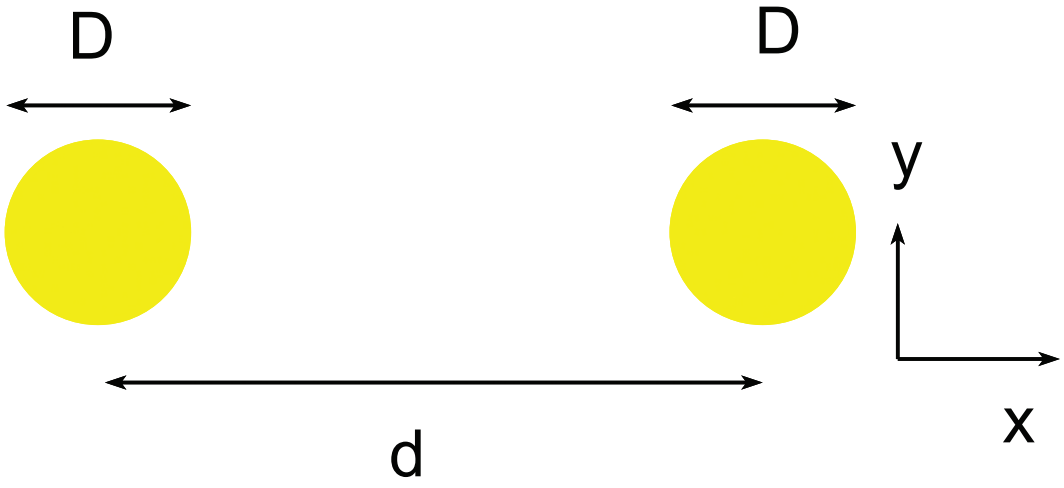

Figure 4.1: Schematic of the system setup. Two nanoparticles with diameter D are separated by a distance d.

the plasmon frequency. This problem can be solved analytically within the quasistatic approximation[19] as shown in the previous chapter, where one ignores the phase retardation and solves the problem using electrostatics. Due to its simplicity, the quasistatic approximation has been widely applied to study systems involving LSPs[21,22,23,24,32] and powerful theoretical methods including transformation optics[33,34] and the hybridization model[23,36] are developed under this approximation. However, by ignoring the phase retardation, one misses many potentially interesting effects arising from multiple scattering.

As in the quantum proximity resonance case, we study such multiple scattering effect in systems consisting of two weakly interacting LSPs, which is popularly known as the dimer system.

The dimer system is being actively explored in the literature due to its richness of fundamental physics and many potential applications . Three regimes have been intensively studied:

1. The Photonic regime[31], where d $\sim$ λ. d, as defined in Fig.4.1, is the center to center distance between the two nanoparticles supporting the LSPs and λ is the wavelength of

the incident light.

2. The strong coupling regime[34], where $D < d \precsim 2D \ll \lambda$ and D is the diameter of the nanoparticles.

3. The quantum tunneling regime[40], where $d - D \sim 0.1$nm(atomic spacing).

Each regime has different interesting physics. For the photonic regime, the spacing between the dimer is comparable to the light wavelength, so one can potentially control the relative phase between the two nanoparticles. As a matter of fact, it is actually more interesting to consider in this regime what happens when one arrange such nanoparticles into a periodic lattice. In that case, a photonic crystal[63] forms and a transmission gap can also form depending on the different lattice spacings.

For the strong coupling field, the two nanoparticles couple through each other's near field and it is no longer valid to assume each other as a individual electric dipole. Since the two gold nanoparticles are so close to each other, they get to influence each other's electric field distribution inside the nanoparticles and we can no longer ignore the fine details of the electric field distribution. In this regime, one can get an antisymmetric mode as well a symmetric mode. The antisymmetric mode starts with a small amplitude, but it increases as the distance between the two nanoparticles decreases.

In the classical limit where the two nanoparticles approach, the amplitude of the symmetric mode can potentially diverge due to the building up of charges of opposite signs on the two sides facing each other, which clearly is not physical. Quantum mechanically, before the two nanoparticles even touch, tunneling will kick in. Tunneling of the electrons from one nanoparticle to the other will eliminate this divergence and prevent the charges from building up further. Beyond this tunneling regime, one also can get a kissing dimer regime where the two nanoparticles do touch each other. This is of course a very different story for both classical

theory and quantum theory. Classically, this kissing dimer case could be handled using the transformation optics[34].

4.3 Intermediate Regime for the Dimer System

In this section, we propose a new simplifying regime for this dimer system, the intermediate regime, which arises when $2D < d < 0.1\lambda$.

This regime was previously named the weak coupling regime in Ref.[34] where it was argued (within the quasistatic approximation) that the dimer system should exhibit the same behavior as individual LSPs. This implies that only a single symmetric mode can be observed in a light scattering experiment.

This quasistatic argument would be correct if the two nanoparticles are non-resonant scatterers of the incident light. However, the LSPs are by nature resonant scattering modes[19] with scattering cross section scaling as λ^2 [19,37]. Thus, the scattering cross section can be orders of magnitude larger than the physical cross section of the nanoparticles. To get a sense of how large this difference is, we consider a typical example involving gold nanoparticles.

For gold nanoparticles, the resonant wavelength is around 540nm. For nanoparticles with D<20nm, the scattering cross section can be 1000 times larger than their physical sizes[1]. If two such nanoparticles are placed within one wavelength, one can expect multiple scattering to yield new interesting physics in certain regimes. In quantum scattering theory, multiple scattering between two resonant scatterers gives rise to the proximity resonance[37]. Due to the distinctive resonant properties of LSPs, multiple scattering displays a different signature in this plasmonic dimer system[34].

In the following, we first consider the dimer system illuminated by an incoming plane wave polarized in the y direction and propagating in the x direction. For nanoparticles with $D < 20nm$, the LSPs can be well described by two resonant dipoles, $\vec{p}_1(\vec{r}_1), \vec{p}_2(\vec{r}_2)$ [19,34,41], where

$\vec{r}_1, \vec{r}_2$ are the positions of the two nanoparticles. These two dipoles have to satisfy the following self-consistent equations:

$$\begin{aligned} \vec{p}_1(\vec{r}_1) &= \alpha(\omega)[\vec{E}_0(\vec{r}_1) + G(\vec{r}_1 - \vec{r}_2)\vec{p}_2(\vec{r}_2)] \\ \vec{p}_2(\vec{r}_2) &= \alpha(\omega)[\vec{E}_0(\vec{r}_2) + G(\vec{r}_2 - \vec{r}_1)\vec{p}_1(\vec{r}_1)] \end{aligned}, \tag{4.1}$$

where $\alpha(\omega)$ is the dynamic electric polarizability tensor of a single gold nanoparticle, $\vec{E}_0(\vec{r})$ is the incoming wave and $G(\vec{r})$ is the interaction tensor defined as[41,44]

$$G(\vec{r}) = (k^2 + \bigtriangledown\bigtriangledown)\frac{e^{ikr}}{r}, \tag{4.2}$$

and k is the light momentum in free space.

A direct expansion of (4.1) yields

$$\begin{aligned} \vec{p}_1(\vec{r}_1) &= \alpha(\omega)\vec{E}_0(\vec{r}_1) + \alpha(\omega)G(\vec{r}_1 - \vec{r}_2)\alpha(\omega)\vec{E}_0(\vec{r}_2) \\ &+ \alpha(\omega)G(\vec{r}_1 - \vec{r}_2)\alpha(\omega)G(\vec{r}_2 - \vec{r}_1)\alpha(\omega)\vec{E}_0(\vec{r}_1) + \cdots, \end{aligned} \tag{4.3}$$

from which it is clear that $\vec{p}_1(\vec{r}_1)$ includes contributions from both the incoming wave and all the waves that are scattered by the nanoparticle at $\vec{r}_2$ and eventually return to $\vec{r}_1$. These include all the possible multiple scattering paths between the two nanoparticles.

For our setup, the solutions to (4.1) can be written as the sum of a symmetric part and an antisymmetric part:

$$\begin{aligned} \vec{p}_s &= \frac{1}{2}\frac{\alpha_{yy}(\omega)[\vec{E}_0(\vec{r}_1) + \vec{E}_0(\vec{r}_2)]}{1 - \alpha_{yy}(\omega)G_{yy}(\vec{r}_1 - \vec{r}_2)} \\ \vec{p}_{as} &= \frac{1}{2}\frac{\alpha_{yy}(\omega)[\vec{E}_0(\vec{r}_1) - \vec{E}_0(\vec{r}_2)]}{1 + \alpha_{yy}(\omega)G_{yy}(\vec{r}_2 - \vec{r}_1)} \end{aligned}, \tag{4.4}$$

where $\alpha_{yy}(\omega)$ and $G_{yy}(\vec{r})$ are the diagonal elements corresponding to the y directions, of the polarizability tensor and the interaction tensor respectively.

With this definition, the solutions are given by

$$\vec{p}_1(\vec{r}_1) = \vec{p}_s + \vec{p}_{as}, \;\; \vec{p}_2(\vec{r}_2) = \vec{p}_s - \vec{p}_{as}. \tag{4.5}$$

As is clear from (4.4), new resonant modes arise when $Re[\alpha_{yy}(\omega)G_{yy}(\vec{r}_2 - \vec{r}_1)] = \pm 1$, corresponding to a symmetric and an antisymmetric mode respectively. The interpretation of such modes is straightforward. When $Re[\alpha_{yy}(\omega)G_{yy}(\vec{r}_2 - \vec{r}_1)] = 1$, the two dipole moments tend to enhance each other in the same direction and the multiple scattering of the wave between them helps build up a resonant mode where the two dipole moments are polarized in the same direction. On the contrary, the two dipole moments tend to enhance each other in the opposite direction when $Re[\alpha_{yy}(\omega)G_{yy}(\vec{r}_2 - \vec{r}_1)] = -1$, which in turn leads to a resonant mode where the two dipoles are oriented in the opposite directions. This is better illustrated in Figure 4.2(b) and (c).

The excitation mechanism of these two resonant modes is different from those in the photonic regime and the strong coupling regime. In the photonic regime, an antisymmetric mode could arise if $d = \lambda/2$, which corresponds to a phase matching condition. However, the antisymmetric mode in this intermediate regime arises when $d < 0.1\lambda$. In this case, the phase accumulation due to multiple scattering enables the excitation of the antisymmetric mode for subwavelength separation between the two nanoparticles. In the strong coupling regime, the antisymmetric mode arises as a result of the hybridization of the individual dipole modes[23], which requires the separation d to be smaller than twice of the diameter of the nanoparticles for the measured extinction cross section to be appreciative in a scattering experiment with plane wave incidence[34]. This condition corresponds to $\lambda/d > 27$ for $D = 15nm$[20] and this mode gains strength as this ratio increases[35]. However, the antisymmetric mode in the intermediate regime can arise for $\lambda/d <= 15$ and loses strength as this ratio increases. More importantly, the antisymmetric mode in the strong coupling regime can arise within the qua-

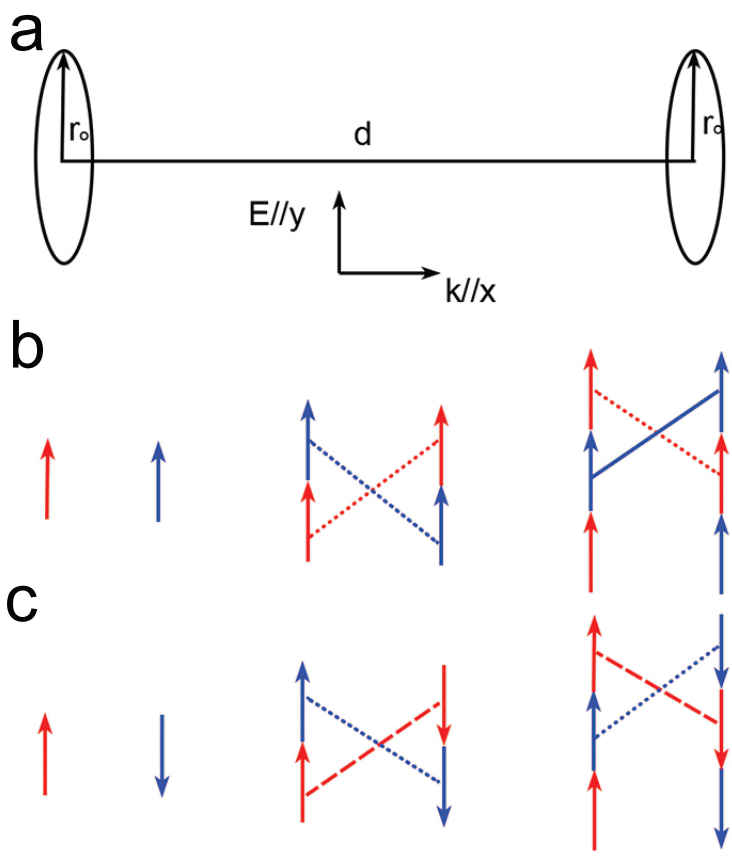

Figure 4.2: Physical Origins of the two modes. (**a**) Schematic of the system setup. Two objects of characteristic length r_o are separated by a distance d. The incoming wave propagates along x and is polarized along y. (**b**) Building up of the symmetric resonant mode. When $Re[\alpha_{yy}(\omega)G_{yy}(\vec{r}_2 - \vec{r}_1)] = 1$, the dipoles tend to enhance each other in the same direction. Starting with two dipoles in the same direction, the red dipole on the left induces an extra red dipole on top of the original blue dipole on the right, as indicated by the red dotted line in the middle. Similarly, the original blue dipole induces an extra blue dipole on top of the original red dipole, indicated by the blue dotted line. These induced dipoles induce further dipoles as shown on the right and the whole process repeats until a resonance is built up. (**c**) Building up of the asymmetric resonant mode. When $Re[\alpha_{yy}(\omega)G_{yy}(\vec{r}_2 - \vec{r}_1)] = -1$, the dipoles tend to enhance each other in the opposite direction. Similar process as in (**b**) leads to an asymmetric resonant mode.

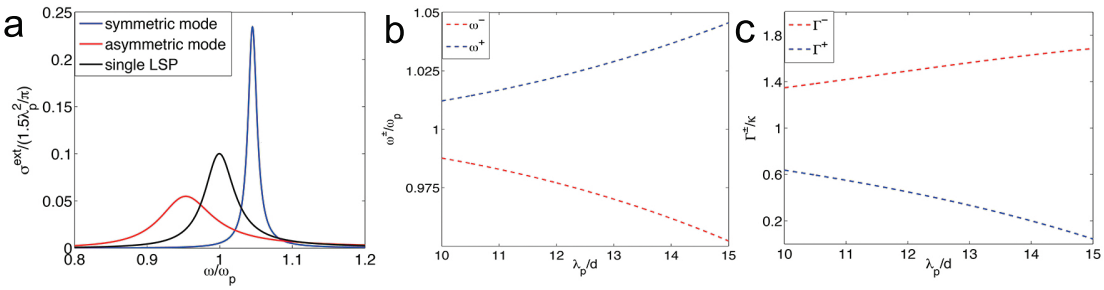

Figure 4.3: (**a**)Extinction cross sections for the single LSP(black), the antisymmetric mode(red), and the symmetric mode(blue). In this plot, λ_p/d is chosen to be 15 and the cross section for the symmetric mode is multiplied by a factor of 0.1. (**b**) The resonance frequencies of the antisymmetric mode and the symmetric mode as a function of λ_p/d. The red dashed line corresponds to the antisymmetric mode while the blue dashed line corresponds to the symmetric mode. The resonance frequencies are measured in the units of the LSP resonance frequency ω_p. (**c**)The damping rates of the antisymmetric mode and the symmetric mode as a function of λ_p/d. The red dashed corresponds to the antisymmetric mode while the blue dashed line corresponds to the symmetric mode. The damping rates are measured in terms of the damping rate of the single LSP damping rate κ. In all three plots, κ_r/κ is fixed to be 10% and ω_p/κ is assumed to be 20.

sistatic approximation[34] where phase retardation is completely ignored. However, as we will demonstrate below, the antisymmetric mode in the intermediate regime won't couple to the incident light in the quasistatic limit, which clearly points to a different origin from that in the strong coupling regime.

For nanoparticles supporting LSPs, $\alpha_{yy}(\omega)$ can be written as

$$\alpha_{yy}(\omega) = \frac{3c^3\kappa_r}{2\omega_p^2}\frac{1}{\omega_p^2 - \omega^2 - i\kappa\omega^3/\omega_p^2}, \tag{4.6}$$

where ω_p is the single LSP resonance frequency, κ is the total damping rate and κ_r is the radiative contribution to κ. This form has the merits of satisfying both the optical theorem and causality in the absence of absorption[38,39].

If we define $\omega_o = \sqrt{3c^3\kappa_r/2\omega_p^2 d^3}$ and keep only the leading terms, we find the following simplified expressions for the resonant frequency and damping rate for the antisymmetric

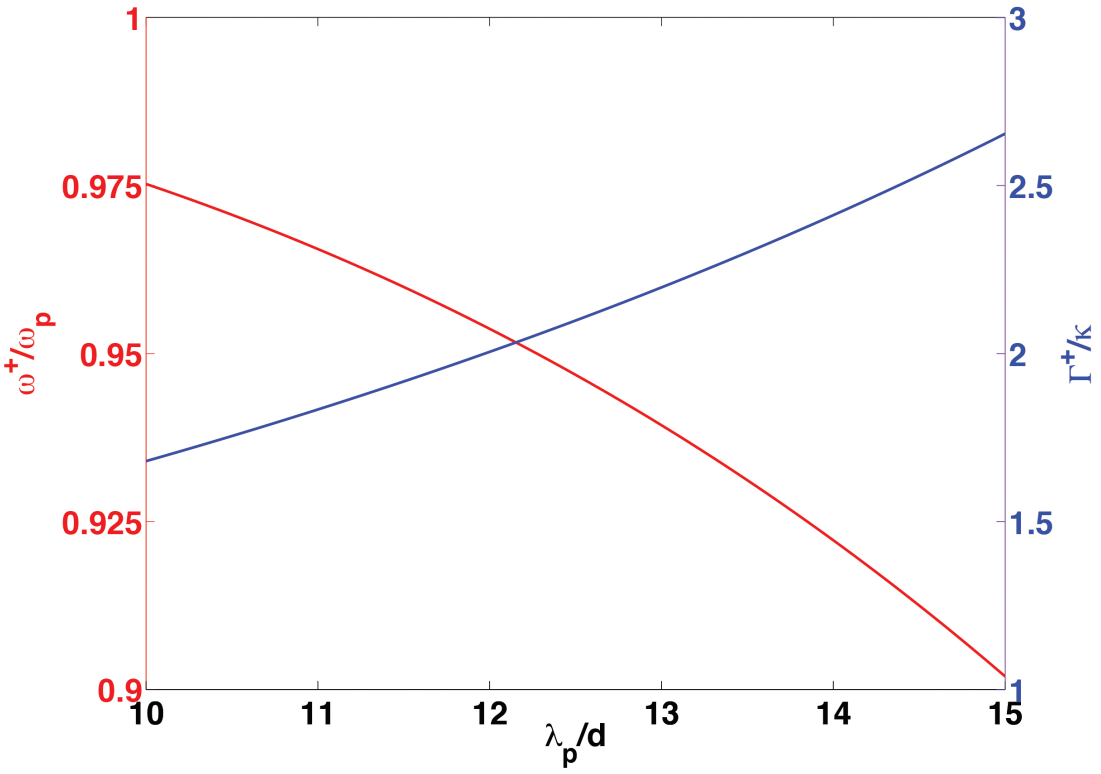

Figure 4.4: Left: The resonance frequency of the symmetric mode as a function of λ_p/d, measured in units of ω_p. Right: The damping rate of the symmetric mode as a function of λ_p/d, measured in units of κ.

mode:

$$\begin{aligned}\omega^- &= \sqrt{\omega_p^2 - \omega_o^2 \cos\frac{\omega_p d}{c}} \\ \Gamma^- &= \frac{\omega_o^2}{\omega^-}\sin\frac{\omega_p d}{c} + \frac{\kappa(\omega^-)^2}{\omega_p^2} \\ &= \kappa + \frac{\omega_o^2}{\omega^-}\sin\frac{\omega_p d}{c} - \frac{\kappa\omega_o^2}{\omega_p^2}\cos\frac{\omega_p d}{c}.\end{aligned} \tag{4.7}$$

For the symmetric mode, the results are

$$\begin{aligned}\omega^+ &= \sqrt{\omega_p^2 + \omega_o^2 \cos\frac{\omega_p d}{c}} \\ \Gamma^+ &= \frac{\kappa(\omega^+)^2}{\omega_p^2} - \frac{\omega_o^2}{\omega^+}\sin\frac{\omega_p d}{c} \\ &= \kappa + \frac{\kappa\omega_o^2}{\omega_p^2}\cos\frac{\omega_p d}{c} - \frac{\omega_o^2}{\omega^+}\sin\frac{\omega_p d}{c}.\end{aligned} \tag{4.8}$$

The above formula implies that, under certain conditions, the damping rate of the antisymmetric mode can increase above the single LSP damping rate κ, while the damping rate of the symmetric mode can drop below κ. These conditions are

$$\begin{aligned}\frac{\omega_p d}{c} &> \frac{\omega^-}{\omega_p^2}\kappa \\ \frac{\omega_p d}{c} &> \frac{\omega^+}{\omega_p^2}\kappa\end{aligned}, \tag{4.9}$$

which can be easily satisfied by gold nanoparticles as shown below.

The total extinction cross section for this system is found to be

$$\sigma^{tot}(\omega) = \sigma^{s}(\omega) + \sigma^{as}(\omega), \tag{4.10}$$

where

$$\sigma^s(\omega) = \frac{8\pi\omega Im[\alpha_{yy}(\omega)]}{c\left|1-\alpha_{yy}(\omega)G_{yy}(\vec{r}_1-\vec{r}_2)\right|^2}$$
$$\sigma^{as}(\omega) = \frac{8\pi\omega Im[\alpha_{yy}(\omega)]}{c\left|1+\alpha_{yy}(\omega)G_{yy}(\vec{r}_1-\vec{r}_2)\right|^2} \quad . \tag{4.11}$$

We define $\sigma^s(\omega)$ as the extinction cross section for the symmetric mode and $\sigma^{as}(\omega)$ as that for the antisymmetric mode. In order to show how effective this multiple scattering mechanism is in exciting the two modes, we plot the two extinction cross sections separately in Fig.4.3(a), using the extinction cross section of the single LSP mode as the reference. Also plotted in Fig.4.3 are the resonance frequencies $\omega^{\pm}/\omega_p$ and the damping rates $\Gamma_{\pm}/\kappa$ as a function of λ_p/d, where $\lambda_p = \omega_p/c$ is the light wavelength at the single LSP frequency. In these calculations, κ_r/κ is fixed to be 10% and ω_p/κ is assumed to be 20, which are chosen based on a previous experiment on gold nanoparticles with $D = 15nm$[20]. When $\lambda_p/d = 15$, the damping rate of the antisymmetric mode is boosted by a factor of 1.8, while the damping rate of the symmetric mode is reduced by a factor of 5. Combined with a 10% ω_p splitting in the resonant frequencies, the two modes can be easily distinguished in an experimental setup.

Even though the antisymmetric mode is weaker than the symmetric mode, it is nonetheless greatly enhanced by multiple scattering. As one can see from Fig.4.3a, the resonance peak for the antisymmetric mode is only smaller than the single LSP resonance peak by a factor of about two. Since the single LSP resonance is strong enough to enable single molecule detection[46], the antisymmetric mode in this intermediate regime can certainly lead to observable effects.

One way to understand the strange behavior of the linewidths of the two resonant modes in this regime is by thinking about what a small or large linewidth means in this case. Taking the symmetric mode for instance, the symmetric mode is calculated to have a sharpened resonance when compared with the single nanoparticle case. What this reduced linewidth means is that if

one changes the incident light frequency to be slightly away from the resonance condition, the symmetric resonance dies off much faster than in the case with the same amount of frequency change for a single nanoparticle resonance. This is because of the asymmetry in the expression for the dynamic polarizability (4.6). For higher frequency/energy, the single nanoparticle resonance is more quickly damped than that for a lower frequency/energy, which makes intuitive sense. For the symmetric mode, the resonance frequency is shifted up when compared with the single LSP resonance due to dipole-dipole interaction. This increased resonance frequency means that the polarizability for each individual nanoparticle at the new resonance frequency is now more sensitive to frequency changes than in the single nanoparticle case. The symmetric mode is excited essentially through a phase matching condition(two dipoles polarizing each other in the same direction). Compared with the case of the antisymmetric mode, which has a reduced resonance frequency, the polarizability changes more in the symmetric mode case for the same amount of frequency change, destroying the phase matching condition much faster than that for the antisymmetric mode. This explains why the linewidth of the symmetric mode should be larger than that of the antisymmetric mode and the single LSP mode.

As mentioned above, what distinguishes this antisymmetric mode in the intermediate regime from that in the strong coupling regime is its reliance on the phase retardation in the incident field. If the phase retardation is removed by changing the polarization of the incident light($\vec{E}_0//x$ and $\vec{k}//y$), the antisymmetric mode is completely suppressed and only the symmetric mode can be excited. The resonance frequency and the damping rate for the symmetric mode in this case are found to be The real and imaginary parts of the resonance frequency for

the symmetric mode are found to be

$$\begin{aligned}\omega'^{+} &= \sqrt{\omega_p^2 - 2\omega_o^2 \cos\frac{\omega_p d}{c}} \\ \Gamma'^{+} &= \frac{\kappa(\omega'^{+})^2}{\omega_p^2} + \frac{2\omega_o^2}{\omega'^{+}} \sin\frac{\omega_p d}{c} \\ &= \kappa - \frac{2\kappa\omega_o^2}{\omega_p^2} \cos\frac{\omega_p d}{c} + \frac{2\omega_o^2}{\omega'^{+}} \sin\frac{\omega_p d}{c}.\end{aligned} \tag{4.12}$$

These relations are plotted in Figure 4.4 using the same parameters as before. In this case, the symmetric mode instead displays enhanced damping.

4.4 Plasmon Wavguiding

The integration and miniaturization of modern electronic devices is rapidly approaching its fundamental limit. One proposal is to replace the electronic signals with optical signals[56]. However, diffraction prohibits the localization of electromagnetic waves below one wavelength, limiting the minimum size of photonic devices to be 1000 times larger than that of electronic devices[57]. Plasmonics, which merges photonics and electronics at the nanoscale, has emerged as a promising solution to this dilemma[56,57,59,19].

One fundamental problem in plasmonics is how to transfer the electromagnetic energy at the nanoscale with both low loss and large field compression. Existing proposals are either based on Surface Plasmon Polaritons(SPPs) bound at the metal-dielectric interfaces[60,61,62,63], or the resonant modes in chains of Localized Surface Plasmons(LSPs) supported by metallic nanostructures[64,65,66,67]. However, a basic trade-off exists between field localization and propagation loss for SPPs. Long-range SPP modes experience only small dissipation, but have poor subwavelength localization, while short-range SPP modes have large dissipation but strong field localization[56,57]. Similarly, the resonant modes in chains of LSPs can compress the electromagnetic field into nanoscale regions, but also exhibit very strong dissipation[66,67].

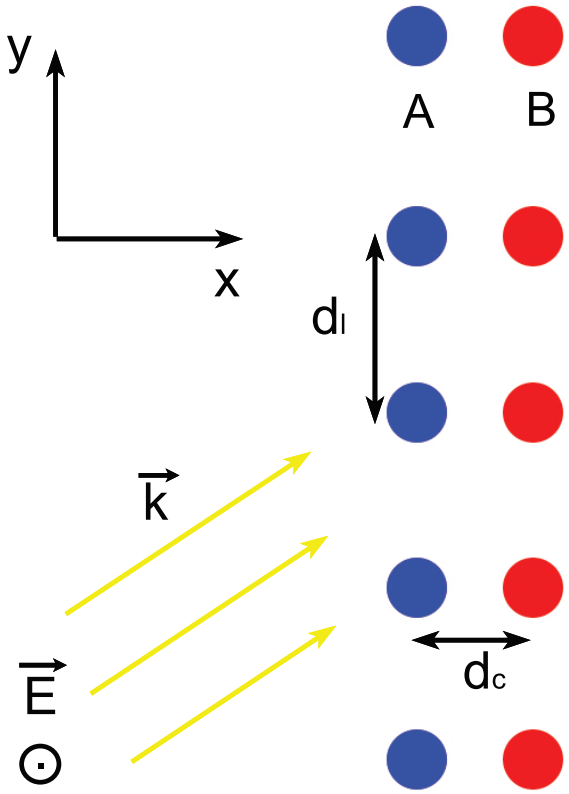

Figure 4.5: Schematic of the system setup for plasmon waveguiding.The nanoparticles are arranged in an infinite one-dimensional linear chain in the x-y plane, with two nanoparticles per unit cell(labeled as A(blue) and B(red)). The lattice spacing is d_l and the inter-particle distance within each unit cell is d_c, both of which are significantly smaller than the wavelength of the incident light, which is linearly polarized in the z direction and propagating in the x-y plane with wavevector $\vec{k}$.

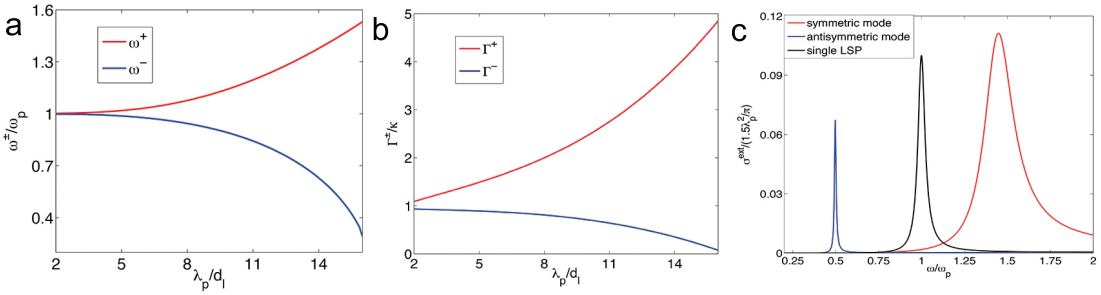

Figure 4.6: Dispersion relation and damping rates for the infinite chain.**(a)** Resonance frequencies for the symmetric mode(ω^+) and the antisymmetric mode(ω^-) as a function of λ_p/d_l. **(b)** Damping rates for the symmetric mode(Γ^+) and the antisymmetric mode(Γ^-) as a function of λ_p/d_l. **(c)** The extinction cross section for each mode when $\lambda_p/d_l = 15$. The extinction cross section for the symmetric mode is normalized for five unit cells, while the extinction cross section for the antisymmetric mode is for fifty unit cells. In all the calculations, the following ratios are used: $\kappa_r/\kappa = 0.1$, $\kappa/\omega_p = 1/20, d_l/d_c = 2$ and $k_y/k_x = 1$.

In this section, we propose an alternative structure based on LSPs and show that a new resonant mode arises with both large field compression and weak dissipation. This mode exists in an intermediate regime where the multiple scattering of the electromagnetic waves between the LSPs has the dominating effect, and is beyond the prediction of the quasi-static approximation, which was previously shown to be able to explain the resonant modes in arrays of graphene microdisks[24]. Even though the strong coupling regime[23,34] and the quantum tunneling regime[40] have been extensively investigated in the literature, this intermediate regime, to our best knowledge, is less explored and deserves more attention.

The setup of our system is shown in Fig.4.5. Nanoparticles supporting the LSPs are arranged in a linear periodic chain in the x-y plane, two per unit cell. The lattice spacing is d_l and the inter-particle distance within each unit cell is d_c, both of which are assumed to be significantly smaller than the wavelength of the incoming light. The incoming light is a plane wave linearly polarized in the z direction and propagating in the x-y plane with wavevector $(k_x, k_y, 0)$. We first consider an infinite chain and numerical results involving a finite chain follow.

We index each unit cell by an integer q and the two nanoparticles in each unit cell are la-

beled either as A(blue in Fig.4.5) or B(red in Fig.4.5). One can then write the total electric field at any position $\vec{r}$ in terms of the contribution from each nanoparticle:

$$\vec{E}(\vec{r}) = \vec{E}_0(\vec{r}) + \sum_{q=-\infty}^{\infty} \sum_{i=A,B} G_{zz}(\vec{r} - \vec{r}_{q,i}) \alpha_{zz}(\omega) \vec{E}(\vec{r}_{q,i}), \tag{4.13}$$

where $\vec{r}_{q,i}$ is the location of the nanoparticle i(A or B) in unit cell q. Each $\vec{E}(\vec{r}_{q,i})$ satisfies a self-consistent condition:

$$\vec{E}(\vec{r}_{q,i}) = \vec{E}_0(\vec{r}_{q,i}) + \sum_{(p,j)\neq(q,i)} G_{zz}(\vec{r}_{q,i} - \vec{r}_{p,j}) \alpha_{zz}(\omega) \vec{E}(\vec{r}_{p,j}). \tag{4.14}$$

In theory, one has such an equation for every pair of (q, i) and those equations need to be solved simultaneously. However, the solution can be significantly simplified if the incoming wave is translation invariant[41]. When $\vec{E}_0(\vec{r}) = E_0 e^{i(k_x x + k_y y)} \hat{z}$, we can relate $\vec{E}(\vec{r}_{q,i})$ to $\vec{E}(\vec{r}_{0,i})$ by

$$\vec{E}(\vec{r}_{q,i}) = \vec{E}(\vec{r}_{0,i}) e^{iqk_y d_l}, \tag{4.15}$$

where d_l is the lattice spacing as defined in Fig.4.

Plugging (4.15) into (4.14), we are left with only two equations:

$$\begin{aligned} \vec{E}(\vec{r}_{0,A}) &= \vec{E}_0(\vec{r}_{0,A}) + T_{A'\to A}\vec{E}(\vec{r}_{0,A}) + T_{B\to A}\vec{E}(\vec{r}_{0,B}) \\ \vec{E}(\vec{r}_{0,B}) &= \vec{E}_0(\vec{r}_{0,B}) + T_{B'\to B}\vec{E}(\vec{r}_{0,B}) + T_{A\to B}\vec{E}(\vec{r}_{0,B}), \end{aligned} \tag{4.16}$$

where $T_{A'\to A}(T_{B'\to B})$ describes the interaction between the A(B) nanoparticle in the unit cell zero and all the other A(B) nanoparticles, and $T_{B\to A}(T_{A\to B})$ describes the interaction between the A(B) nanoparticle in the unit cell zero and all the B(A) nanoparticles. These four terms

summarize all the possible multiple scattering effects in this system and they are given by

$$\begin{aligned} T_{A'\to A} &= \sum_{p\neq 0} G_{zz}(\vec{r}_{0,A}-\vec{r}_{p,A})\alpha_{zz}(\omega)e^{ik_y pd_l} \\ T_{B'\to B} &= \sum_{p\neq 0} G_{zz}(\vec{r}_{0,B}-\vec{r}_{p,B})\alpha_{zz}(\omega)e^{ik_y pd_l} \\ T_{A\to B} &= \sum_{p=-\infty}^{\infty} G_{zz}(\vec{r}_{0,B}-\vec{r}_{p,A})\alpha_{zz}(\omega)e^{ik_y pd_l} \\ T_{B\to A} &= \sum_{p=-\infty}^{\infty} G_{zz}(\vec{r}_{0,A}-\vec{r}_{p,B})\alpha_{zz}(\omega)e^{ik_y pd_l} \end{aligned} . \tag{4.17}$$

For our system, $T_{A'\to A} = T_{B'\to B}$, $T_{A\to B} = T_{B\to A}$, and the solution to (5.3) is

$$\begin{aligned} \vec{E}(\vec{r}_{0,A}) &= \frac{1}{2}\left(\frac{\vec{E}(\vec{r}_{0,A})+\vec{E}(\vec{r}_{0,B})}{1-T_{A'\to A}-T_{B\to A}} + \frac{\vec{E}(\vec{r}_{0,A})-\vec{E}(\vec{r}_{0,B})}{1-T_{A'\to A}+T_{B\to A}}\right) \\ \vec{E}(\vec{r}_{0,B}) &= \frac{1}{2}\left(\frac{\vec{E}(\vec{r}_{0,A})+\vec{E}(\vec{r}_{0,B})}{1-T_{A'\to A}-T_{B\to A}} - \frac{\vec{E}(\vec{r}_{0,A})-\vec{E}(\vec{r}_{0,B})}{1-T_{A'\to A}+T_{B\to A}}\right) \end{aligned} . \tag{4.18}$$

Resonance arises when either one of the following conditions is satisfied

$$\begin{aligned} Re[T_{A'\to A}+T_{B\to A}] &= 1 \\ Re[T_{A'\to A}-T_{B\to A}] &= 1 \end{aligned} , \tag{4.19}$$

corresponding to a symmetric mode and an antisymmetric mode respectively.

If $Re[T_{A'\to A}+T_{B\to A}] = 1, \vec{E}(\vec{r}_{0,A}) = \vec{E}(\vec{r}_{0,B})$, corresponding to a symmetric mode where the two LSPs in each unit cell are polarized in the same direction, and it is an enhanced dipole mode. If $Re[T_{A'\to A}-T_{B\to A}] = 1, \vec{E}(\vec{r}_{0,A}) = -\vec{E}(\vec{r}_{0,B})$ and it corresponds to an antisymmetric mode, where the two LSPs in each unit cell are polarized in the opposite directions. This is better seen in Fig.4.7, where we plot the electric field distribution for each mode. For the symmetric mode, the real part of the electric field is enhanced in the middle of each unit cell, while a minimum is instead achieved for the antisymmetric mode . For both modes, the

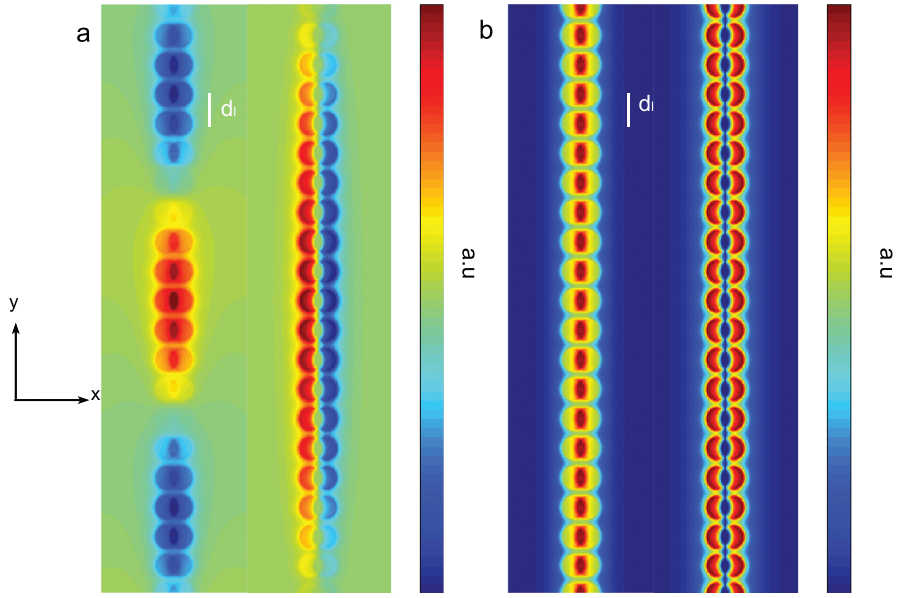

Figure 4.7: Electric Field Distribution for the infinite chain.**(a)** Real part of the electric field E_z in the z=0 plane for each mode in the infinite chain. Left: the symmetric mode. Right: the antisymmetric mode. **(b)**The amplitude $|E_z|$ in the z=0 plane for the infinite chain. Left: the symmetric mode. Right: the antisymmetric mode. In all the simulations, the following ratios are used: $\kappa_r/\kappa = 0.1, \kappa/\omega_p = 1/20, \lambda_p/d_l = 15, d_l/d_c = 2$ and $k_y/k_x = 1$.

amplitude of the electric field is a periodic function in y with period d_l. From the real part of the electric field, we can also infer that the antisymmetric mode has a longer wavelength and smaller resonance frequency.

The resonance frequencies can be found by solving (4.19) numerically and they depend only on the ratios of λ_p/d_l, d_l/d_c and k_y/k_x. Once we find the resonance frequencies, we can plug them back into (4.18) and find the corresponding damping rates and extinction cross sections. These results are shown in Fig.4.6.

Unlike the dimer case, the antisymmetric mode has a significantly reduced damping rate(by a factor up to ten), while the symmetric mode is strongly damped. This reversal in behavior is due to the competition between the short-range near field coupling and the long-range radiative coupling. For the dimer case, the near field coupling dominates while the long-range radiative coupling dominates for the infinite chain.

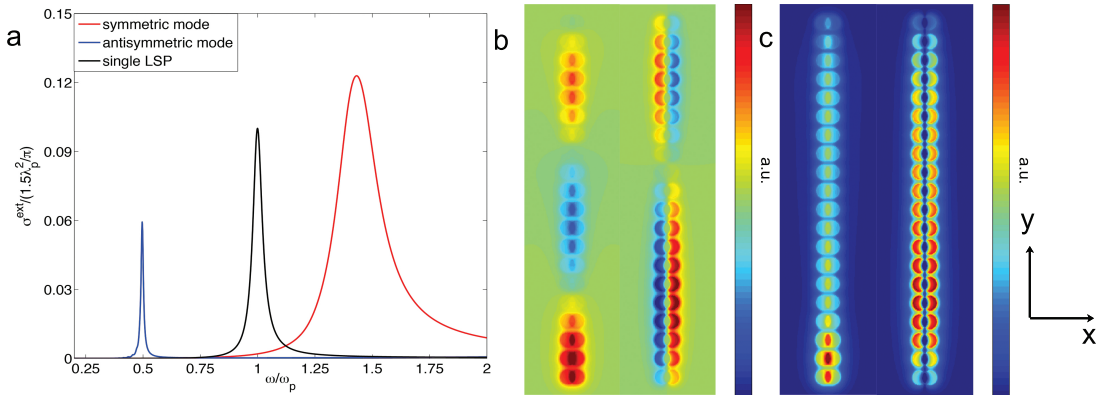

Figure 4.8: Dispersion relation and damping rates for the finite chain.**(a)** Extinction cross sections for a finite chain consisting of 20 unit cells. The extinction cross section for the antisymmetric mode is normalized for fifty unit cells, while the cross section for the symmetric mode is normalized for five unit cells for comparison with Fig.4.6. **(b)** Real part of E_z for the symmetric mode(left) and the antisymmetric mode(right). (c) The amplitude $|E_z|$ for the symmetric mode(left) and the antisymmetric mode(right). In all the calculations, the following ratios are used: $\lambda_p/d_l = 15, \kappa_r/\kappa = 0.1$, $\kappa/\omega_p = 1/20, d_l/d_c = 2$ and $k_y/k_x = 1$.

If we define a compression ratio as λ/d_l, where λ is the light wavelength at the resonance frequency, it can reach up to 30 for the antisymmetric mode. Thus, this antisymmetric mode has the merits of both very weak dissipation and large field compression.

We plot the electric field distribution in Fig.4.7. As expected, the field distribution is periodic and one can easily spot the difference between the symmetric mode and antisymmetric mode.

In all the above calculations, we have assumed the following ratios: $\kappa_r/\kappa = 0.1, \kappa/\omega_p = 1/20$, $d_l/d_c = 2$ and $k_y/k_x = 1$. These values are chosen to match the experimental measurements on gold nanorods with aspect ratio 3:1 and radius 7.5nm[20].

For finite chains, translational symmetry is broken and (4.15) no longer holds. In this case, the boundary could have a significant influence on the excited modes.

To study this finite case, we numerically simulate a chain consisting of only twenty unit cells using the method in Ref.[58] and the results are presented in Fig.4.8.

As we can see from Fig.4.8, the antisymmetric mode remains a sharp resonance with a com-

pression ratio approaching 30 and its damping rate is reduced by a factor of five. The electric field distribution for each mode is plotted in Fig.4.8(b) and (c).

It is worth mentioning that the reversal in behavior of the two resonant modes in terms of linewidths when compared with the previous case is due to the long range radiative coupling in the chained system, which has the dominant effect, while only near field matters in the plasmon dimer case.

4.5 Summary

To summarize, we introduce a new simplifying intermediate regime for studying the plasmon dimer system, where multiple scattering has the dominant effect. We show that one can excite the antisymmetric mode by direct light illumination. It was previously believed that this antisymmetric mode, important for plasmon-induced transparency[49] and three-dimensional plasmon rulers[50], can't be excited by the incident light in this "weak coupling" regime and one has to resort to either electron beam[45] or designed structures with broken symmetry[47,48,49,50] to activate it. More importantly, we show that multiple scattering can lead to significant changes in the damping rates of the plasmon modes in this new regime. The damping rate of the symmetric mode, a highly radiating mode, can be reduced by a factor of five, while the damping rate of the antisymmetric mode, a subradiant mode, is enhanced above the single LSP damping rate. The opposite behavior is observed for both the infinite and finite one dimensional chains of LSPs with two LSPs per unit cell. In this case, multiple scattering and the long-range radiative coupling give rise to an antisymmetric mode with both subwavelength field compression and low dissipation. This could be important for many practical applications, such as Surface-Enhanced Raman Scattering[51], three-dimensional plasmon rulers[50], improved photovoltaic devices[5], ultrasensitive biosensing[27], optical emitters[35,26] and plasmon waveguides[52]. But most importantly, manipulating plasmon modes and their damping rate are of key

importance in the field of plasmonics, and our results provide a new route to explore to this end.

5

Putting the Electromagnetic Fields into Shapes

The ability to manipulate electromagnetic fields on the subwavelength scale could offer unprecedented opportunities in many fields of science, ranging from ultrasensitive biosensing[27], data processing[38] to new platforms for probing many-body physics[40]. Enabled by the great freedom of design provided by artificially engineered metamaterials[70,71,72], transformation optics[73,74] offers a way to obtain control over electromagnetic fields through locally varying a material's permittivity and permeability, leading to exciting functionalities such as cloaking[71,75], optical illusion[77] and metamaterial analog computing[79].

Here we demonstrate another possibility for manipulating electromagnetic fields, without the need to engineer a material's electromagnetic properties. Our design takes advantage of the enormous size of the configuration space in combinatorial problems[80] and the resonant scattering properties of metallic nanoparticles, which, when illuminated by light with the right frequency, give rise to resonant scattering modes known as the Localized Surface Plasmons(LSPs)[19]. These LSP modes bridge the gap between photonics and electronics[56,57], and in many ways behave like the atomic/molecular point scatterers used to build quantum corrals[81,82]. Our design does not require the spatial engineering of a material's electromagnetic properties and can be made effective in the whole visible spectrum. Moreover, one can easily extend the approach to other spectra by replacing the metallic nanoparticles with other scatterers of light.

5.1 Quantum Corral

Quantum corral is one of the most successful products of the Scanning Tunneling Microscopy(STM) technology. In a nutshell, quantum corral forms by confining electron surface wave inside a corral formed by densely populated scatterers.

The first set of experiment probing the electron wave pattern insider a quantum corral was reported in 1993[83] and the first multiple scattering theory explaining the formation of quantum corral was proposed by Prof. Heller in the following year[84]. In the experiment[83], iron atoms were moved by a STM tip onto a Cu(111) surface. The iron atoms are densely spaced around the perimeter of a circle and these iron atoms serve as the scatterers of the electrons underneath inside the copper substrate.

On the copper substrate exists a type of surface states known as the Shockley surface states and these are the only electron states available for scattering at the Fermi energy. In a typical experimental setup[81], the wavelength of these surface states is around 29.5 $\overset{o}{A}$, which is

much larger than the size of the scatterers, iron atoms. When the physical size of the scatterer is much larger than the wavelength of the electrons, one can make the s-wave scattering approximation.

In the experiment, the STM tip is raster scanning across the surface and recording the signal measured in each position. This pattern was observed to be a standing wave pattern in the initial experiment[83] and the authors used an "electron in the box" model to explain the formation of the standing wave pattern. This explanation works reasonably well when the iron atoms are densely packed around the perimeter, but it does not account for the wave pattern outside the box. In the subsequent year, a multiple scattering theory based on s-wave scattering was proposed to explain both the observed standing wave pattern , resonance energies and the linewidth[83]. In the multiple scattering approach, a theoretical equivalence is established between the Local Density of States(LDOS), the measurement in the STM, and the backscattered amplitude, the quantity calculated using the theory.

5.2 Optical Corral

For quantum corrals, the multiple scattering of electron surface waves by the point scatterers leads to "standing" wave patterns that could be probed by a scanning tunneling microscope(STM). The optical analogy of quantum corral was theoretically predicted[86] and experimentally probed using a scanning near field optical microscope[87].

For the optical corral, metallic nanoparticles are used as scatterers of the incident light. As discussed in the previous chapter, the size of each metallic nanoparticle is much smaller than the wavelength of the incident light. Thus, each nanoparticle responds to the incident light very much as a point electric dipole moment. This makes the physics very much similar to the quantum corral case, where each iron atom serves a point s-wave scatterer. In both case, the scatterer's size is much smaller than the wavelength of the wave being scattered and the way

they scatter the incident wave involves only one mode, which makes the calculation much easier.

For the optical corral, the metallic nanoparticles are arranged on the surface of a two dimensional dielectric substrate. Given their configuration on the two dimensional surface, one can solve a simple forward scattering problem[86] to find out the resulting measured wave pattern. Different configurations of nanoparticles lead to different wave patterns. In our project, we answer the question:" Given a desired wave pattern, can we find a way to arrange the metallic particles that will yield this pattern?"

5.3 Scattering-type Near-field Optical Microscopy

Before we get into the details of the setup for solving the inverse problem, we will first give a brief introduction to the experimental technique for probing the electromagnetic field at the nanoscale.

The state of the art technique is called scattering-type Near-field Optical Microscopy(s-SNOM) and it was reported that the best spatial resolution it could achieve is 1nm[85].

The setup is shown in Fig.5.1. A metallic tip is held on top of a surface being probed and this tip is at the same time illuminated by a focused laser beam. The measurement in the experiment is the amplitude being backscattered towards the tip. By recording the signal measured at each position, this can provide information about the electromagnetic properties of the two dimensional surface.

The metallic tip is usually modeled as an electric dipole polarized in the z direction and it sends out light in all direction. The response of the dielectric substrate is modeled within the quasi-static approximation, since the distance between the tip and the surface is usually orders of magnitude smaller than the wavelength of the light. That is, the substrate responds to the electric dipole by generating a mirror point dipole with a polarizability equal to $\alpha\beta$, where α is

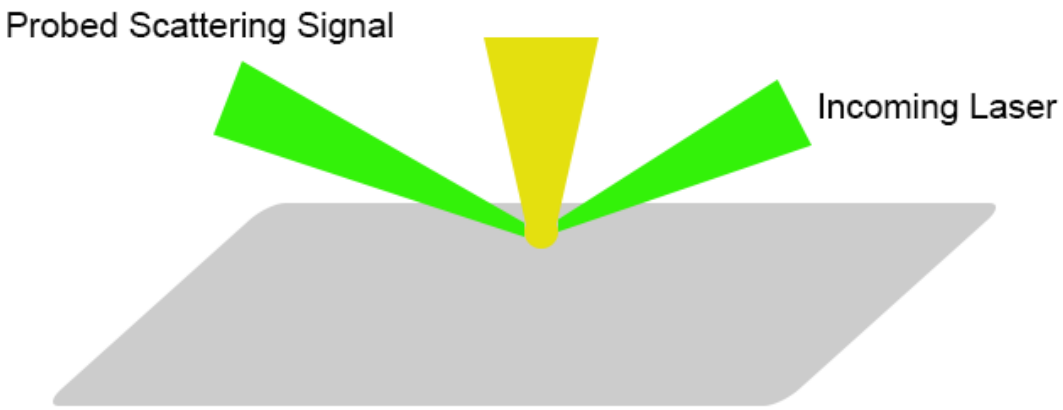

Figure 5.1: A focused light beam illuminates a metallic tip , which is held on top of a two dimensional surface to be probed. The scattered light is collected and brought to interfere with a reference laser beam to determine its amplitude and phase. By moving the tip around and recording the measured signal, one gets a two dimensional map of the field distribution.

the polarizability of the tip and β is

$$\beta = \frac{\epsilon_s - 1}{\epsilon_s + 1}, \tag{5.1}$$

where ϵ_s is the dielectric constant of the substrate[88,89].

Overall, the effective polarizability of the metallic tip together with the substrate is

$$\alpha_{eff} = \frac{\alpha(1+\beta)}{1 - \alpha\beta/(16\pi(a+z)^3)}, \tag{5.2}$$

where a is the radius of the tip, which is modeled as a metallic sphere and z is the distance from the center of the tip to the surface of the substrate.

Thus, one can just assume that overall electric response of the tip and the substrate is to generate an effective electric dipole in the z direction. When there are additional structure existent on the surface, some of the wave will be backscattered towards the tip and the backscattered wave will be interfering with the incoming field and also contributes to the signal. The collected total backscattered wave is then brought into interference with a reference beam. The

interference pattern is used to determine both the phase and amplitude of the backscattered wave. In the experiment, the second harmonics is usually measured instead to reduce the noise level from background scattering[88,89].

By moving the tip across the two dimensional surface and recording the corresponding backscattered wave, one can thus obtain a two dimensional image of the local electric response of the material.

At first sight, it seems that s-SNOM operates based on very different physical principles when compared with STM: STM is based on electron tunneling, but s-SNOM is based on light scattering. However, if one takes a multiple scattering perspective, the two stories are actually quite similar in flavor.

According to the multiple scattering understanding of the quantum corrals[84], the STM tip can be understood as an electron source that emits electrons in all direction. The emitted electron then moves freely on the two dimensional surface until it is scattered by some scatterers. Some of the scattered electron wave will be directed back towards the tip and the interference between the backscattered electron wave and the incident wave constitutes the signal being measured by the STM tip.

Similarly in the case of s-SNOM, the metallic tip acts as an electric dipole source that sends light in all directions. The light moves free in the free space until it gets scattered by whatever obstruction there is on the surface. Some of the light will be backscattered towards the tip itself and will interfere with the incoming field. This interference amplitude is being recorded and constitutes the signal being probed.

Thus, according to this multiple scattering formulation, both STM and s-SNOM can be understood within the same physical picture.

5.4 Approach for shaping electromagnetic fields

Our approach for shaping electromagnetic fields relies on solving an inverse scattering problem where the resulting wave pattern is prespecified and the configuration of nanoparticles leading to that wave pattern needs to be determined.

A simple calculation reveals the enormous size of this configuration space. For 80 gold nanoparticles(each with diameter 20nm) arranged within a two dimensional square of side length $3.4\mu m$(about six light wavelengths at the LSP resonance frequency), the number of all possible configurations is more than 10^{235}. This is even orders of magnitude larger than the total number of atoms in the observable universe.

This enormous configuration space presents both challenges and promise for a possible solution of the inverse problem. In terms of challenges, dealing with such a huge configuration space is not easy. Given limited computing resources, one can only sample a small portion of this configuration space and it is not guaranteed that one can find a given configuration with limited amount of time. On the other hand, since the configuration space is so huge, one has also a high chance of finding any pre-specified pattern.

In the following, we consider two possible schemes for solving this inverse problem and in both cases, the control over the electromagnetic field is obtained through the manipulation of the nanoparticles' positions. Precise control of a nanoparticle's position can be achieved using either a near field optical nanotweezer[99] or Atomic Force Microscope tip[78].

Our first setup closely mimics that for the quantum corral[81,82] and can be thought of as a two dimensional plasmonic hologram. In this setup, gold nanoparticles are arranged on the two dimensional surface of a dielectric substrate and a scattering type near field optical microscope(s-SNOM)[88,89,90,91] tip is raster scanning across the surface while being illuminated by a focused laser beam at the LSP resonance frequency. The tip acts both as an illumination source and a probe that reads out the optical signal.

As the source, the tip sends out light in all directions, which then propagates freely in space until being scattered by the gold nanoparticles. Some of the scattered light will be directed back towards the tip and they constitute the signal being read out. As one moves the tip across the surface and records the signal at every position, it yields a two dimensional wave pattern. It is this wave pattern that we will reverse engineer.

In the second setup, the sample is assumed to be illuminated by a plane wave with the LSP resonance frequency. In this case, we do not necessarily need to probe the electric field distribution on top of the sample(even though one can still use s-SNOM in the probing mode to measure the field distribution if one wants to) and instead we engineer the illumination pattern itself. This setup is more relevant to the design of flexible optical traps.

5.5 Solving the forward problem

Our solution of the inverse problem relies on an efficient method to solve the forward scattering problem, that is, given the positions of the metallic nanoparticles on the substrate, we need to find the resulting pattern being probed by the s-SNOM tip or the illumination pattern(incoming field plus scattered field without the tip).

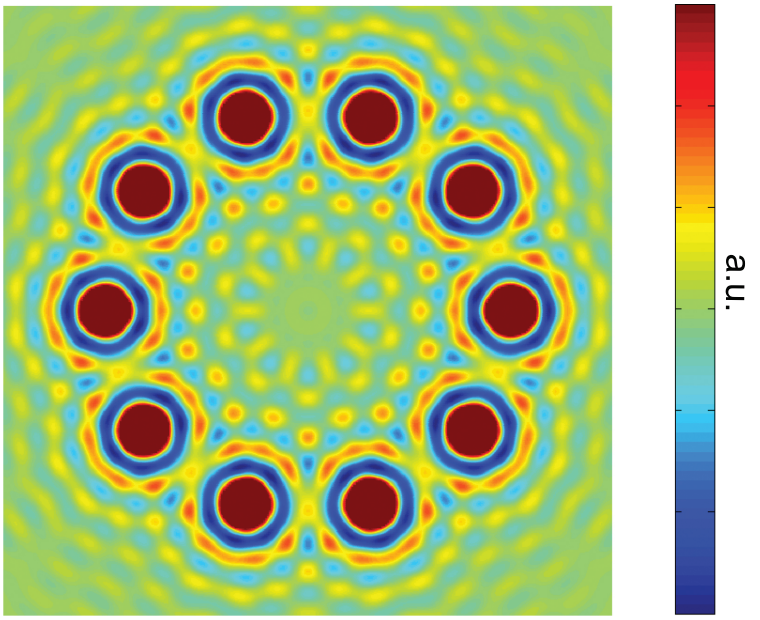

Figure 5.2: Plasmonic Corral.The signal measured by a raster scanning s-SNOM tip in the illumination mode(the tip is polarized in the out of plane direction) when ten nanoparticles are uniformly arranged around the perimeter of a circle with radius 3λ, where λ is the wavelength of the incident light at the LSP resonance frequency.

For this forward scattering problem, we adopt the same multiple scattering approach that successfully explains both the quantum corral[81,82] and the optical corral[86,87]. In this approach, one first solves for the total electric field at each nanoparticle's position using a set of self-consistent equations:

$$\vec{E}(\vec{r}_i) = \vec{E}_0(\vec{r}_i) + \sum_{j \neq i} G(\vec{r}_i, \vec{r}_j)\alpha(\omega)\vec{E}(\vec{r}_j), \tag{5.3}$$

where $\vec{r}_i$ is the position of the ith nanoparticle(out of a total of N nanoparticles). $\vec{E}_0(\vec{r})$ is the incident electric field and $\alpha(\omega)$ is the electric polarizability tensor that characterizes the scattering properties of the nanoparticles.

$G(\vec{r}, \vec{r}')$ is the interaction tensor that describes how the electromagnetic field propagates in the absence of the nanoparticles, and it includes contributions from both free space and the

substrate. Therefore, we can decompose it into two parts:

$$G(\vec{r},\vec{r}') = G_0(\vec{r},\vec{r}') + G_S(\vec{r},\vec{r}'). \tag{5.4}$$

$G_0(\vec{r},\vec{r}')$ is the interaction tensor in free space and it is given by

$$G_0(\vec{r},\vec{r}') = (k^2 + \bigtriangledown\bigtriangledown)\frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}, \tag{5.5}$$

where k is the light momentum in free space.

$G_S(\vec{r},\vec{r}')$ is the interaction tensor generated by the substrate alone and it is provided in Ref.[94]. Alternatively, one can also handle this term within the quasi-static approximation with the methods of image charges[44].

Once we find all the $\vec{E}(\vec{r}_i)$ after solving (5.3), the total electric field at any position $\vec{r}$ can be found using:

$$\vec{E}(\vec{r}) = \vec{E}_0(\vec{r}) + \sum_{i=1}^{N} G(\vec{r},\vec{r}_i)\alpha(\omega)\vec{E}(\vec{r}_i). \tag{5.6}$$

This method is shown to reproduce both the experimental results[87] and more accurate numerical solutions[43,34,41] when the minimum separation between the nanoparticles is larger than twice their diameter, which we impose as a constraint in our algorithm. As an example, we apply this method to calculate the wave pattern probed by a s-SNOM tip in the illumination mode when ten gold nanospheres, each with diameter 20nm, are placed around the perimeter of a circle of radius 3λ, where λ is the incident light wavelength at the LSP resonance frequency.

5.6 Simulated Annealing

Simulated Annealing is one of the best examples where a physical phenomenon can be used to solve very hard optimization problems[80,96].

The best way to describe an annealing process is by giving an example. Suppose now we have a glass full of small ice cubes and our target is to turn this glass of small ice cubes into a whole large ice cube. A glass of small ice cubes is usually considered to be a higher energy state since it is less ordered while a large whole ice cube, given the constraint of the shape of the glass, can be thought as the ground state, which has the lowest energy. So in the words of applied mathematician, we have got an optimization problem to solve here. We are essentially trying to go from a high energy state, a glass of small ice cubes, to the lowest energy state, a whole ice cube. The traditional way to solve an optimization problem is by an iterative approach. In an iterative approach, the configuration space or the parameter space is explored in such a way that the value of the objective function to be minimized decreases monotonically.

However, the way Nature solves this problem is actually kind of interesting, at least quite different from what a traditional applied mathematician would solve it. Instead of sampling the configuration space in ways that monotonically decrease the energy, ”Nature” actually heats up the system to a very high temperature state, which in the example of the ice cubes means that the glass of the small ice cubes got heated such that the ice cubes melt into water. Once we are in the high temperature state, one can then gradually lower the temperature. At each temperature, the system is let to explore the configuration space in such a way that it reaches a thermal equilibrium before going into a lower temperature. As one lowers the temperature, the water molecule’s motion gets limited little by little and it gets to explore a smaller subset of the whole configuration space. Physically, if one lowers the temperature slowly enough, the system will eventually reach the ground state, a whole ice cube. Thus, in this annealing process, Nature solves this complicated optimization problem by allowing con-

trolled uphill energy changes. How many such changes are allowed and how far it can explore depends on the temperature parameter.

Simulated annealing is essentially an algorithm designed to mimic this physical process. In this algorithm, the objective function to be minimized is defined as the energy of the system and we also need to introduce an artificial temperature into the system to tune the number of uphill changes accepted.

The criterion for a thermal equilibrium in the physical sense is that the system reaches a state such that its distribution can be described by a Boltzmann distribution:

$$P(E) \propto exp(-\frac{E}{kT}), \tag{5.7}$$

where k is the Boltzmann constant.

The way a computer can simulate such a thermal equilibrium state is by using a Monte Carlo method known as the Metropolis-Hastings algorithm[98]. It works by generating a Markov Chain whose equilibrium state converges to the Boltzmann distribution.

One can describe a Markov Chain completely by its transition matrix or graph. For a Markov Chain, the state at the next time step only depends on the immediate previous state and is independent of all the other previous state, that is

$$P(x_{t+1}|\{x_i : i < t+1\}) = P(x_{t+1}|x_t) \stackrel{def}{=} P(x_t \to x_{t+1}). \tag{5.8}$$

This transition probability can in general also be time dependent, but for our case, it is independent of time. The idea behind the Metropolis-Hastings algorithm is completely based on detailed balance. Suppose the target equilibrium distribution is $P(x^i)$, where x^i is one of the

possible state in the configuration space, the condition for detailed balance is then given by

$$P(x^i)P(x^i \to x^j) = P(x^j)P(x^j \to x^i). \tag{5.9}$$

In physical terms, (5.9) says that the probability outflow going from state x^i to x^j is the same as the probability inflow going from x^j to x^i in an equilibrium state. If the above condition holds for all possible pairs of i and j, one can make sure that the equilibrium distribution is given by $P(x^i)$. Detailed balance is a sufficient condition to achieve the equilibrium/target distribution, but it is not a necessary condition.

In the design of such a Markov chain, the target distribution is given a prior, up to a normalizing constant. This is actually a very appealing property of this algorithm, since the normalizing constant for high dimensional system or the partition function as in statistical physics is not always known or easy to obtain. Given this unnormalized target distribution, the task is then to obtain a transition matrix that will satisfy the detailed balance condition.

If we go back to the case of a thermal equilibrium, we know that the desired equilibrium distribution is proportional to $exp(-E/T)$ (In the following, we will set $k = 1$.)and one transition probability that satisfies the detailed balance condition is

$$P(E_1 \to E_2) = min(1, exp(-\frac{E_2 - E_1}{T})). \tag{5.10}$$

It is straightforward to verify that

$$\begin{aligned} P(E_1)P(E_1 \to E_2) &= min(exp(-\frac{E_1}{T}), exp(-\frac{E_2}{T})) \\ &= min(exp(-\frac{E_2}{T}), exp(-\frac{E_1}{T})) \\ &= P(E_2)P(E_2 \to E_1). \end{aligned} \tag{5.11}$$

Thus, detailed balance is satisfied and one can simulate the thermal equilibrium state at temperature T by generating a Markov Chain with such transition probability.

One technical detail that is worth pointing out is that in order for a stationary distribution to exist, the Markov Chain needs to be irreducible and positive recurrent. ”Irreducible” means that the states in the Markov Chain are all fully connected, so if one starts at one state, it can get to any other states with positive probability. This would not be possible if there are cycles. The other condition, ”positive recurrent” means that if one starts at one states, there is always a positive non-vanishing probability that it will return to the original state in the long run, no matter how long the Markov Chain has been propagated. This would not be possible if there are absorbing states in the state space. When such a limiting distribution exists, it is also unique.

To summarize, the simulated annealing can be implement in four steps.

The first step is initialization. In this step, one starts with a random initial state and also needs to decide an initial temperature. This temperature can in general be determined by taking the average energy fluctuation or standard deviation of the energy for a series of randomly generated states. The reason behind this is that one wants to start at a high temperature state, where all possible configuration is equally likely and the temperature is just a measure of the underlying energy fluctuations.

The second step is the thermalization step. At each temperature, one needs to establish a thermal equilibrium at that temperature. This is accomplished by using the Metropolis-Hastings algorithm and designing a Markov Chain satisfying the detailed balance condition.

The third step is the cooling step. One needs to come up with a cooling schedule for this step. The cooling schedule is essentially a plan on how one would lower the temperature. In practice, two ways are in general used to decrease the temperature. One way is to decrease the temperature by a fixed amount, usually a small fraction of the initial temperature. The other

way is to decrease the temperature by a fixed percentage. In the first case, the temperature decreases linearly, which can be a little bit dangerous near the phase transition point, where one should spend more time on. In the second case, the temperature is lowered exponentially slowly and the decrease in temperature is getting smaller and smaller. Of course, one needs a ratio near one for this schedule to not miss the phase transition point as well. The second and the third step has to be repeated many times before going to the fourth step.

The final step is the termination step. In this step, one needs a criterion to stop the above two steps. Usually, there are a few heuristic criteria one can use to terminate the process. One of them is to set a target temperature, usually a small fraction of the initial temperature, such that whenever this temperature is reached, the process is terminated. Another way is to set a lower bound on the average energy change. If the average energy decrease falls under a certain value, the process shall be terminated. A third method is to keep track of the fraction of changes accepted at each temperature. If this fraction falls under certain threshold, the process is terminated.

It is worth mentioning that as the temperature gets smaller and smaller, the number of changes accepted will also get smaller and smaller if one stay at each temperature for a fixed amount of steps. However, as true for the physical annealing process as well, one needs to spend more time at lower temperature, especially near the phase transition temperature. One way to do this is to set up a lower bound to the minimum number of changes to be accepted before one moves to another temperature.

Before we conclude on this simulated annealing, we want to mention another point about simulated annealing: Simulated annealing is nothing more than a heuristic algorithm and there is no guarantee that it will ever find the global minimum in a finite number of time. Mathematically, it was shown[96] that in the limit where one lower the temperature infinitely slowly, simulated annealing will converge to the true global minimum. However, the proof usually

relies on the assumption that one reaches a perfect thermal equilibrium at each temperature such that the whole configuration space gets sampled according to the Boltzmann distribution. However, it is far from being a reasonable assumption in our case. If one really achieves a true thermal equilibrium, there is no point in doing a cooling schedule then. One should just start with a low enough temperature and find the equalbrium state at that temperature. In our case, as we mentioned above, the configuration has a size on the order of 10^{235}, but the simulated annealing is finished in about ten millions steps, so one explores almost zero percent of the whole configuration space in the whole annealing process. Thus, it is far from being true to even claim the system is in thermal equilibrium at any temperature.

A better way to think about simulated annealing is to compare it with the iterative approach. As mentioned before, iterative algorithms usually only permit changes in the energy downward direction. Thus, if the function to be minimized is not convex and is multimodal, as in most combinatorial optimization problem, which local minimum is found depends highly on the initial condition. When the objective function or the energy function is multimodal, which it usually is in higher dimensional space, this can be a very significant problem. The simulated annealing is an improvement to this iterative approach in the sense that it does not just allow changes that decrease the energy, but it also allows changes that increase the energy while maintaining an overall downward trend in energy in the process. By accepting changes that will increase the energy, it prevents the system from getting stuck in a shallow minimum, which nevertheless is impossible to escape in the iterative approach.

5.7 Genetic Algorithms

Genetic algorithms[97] are another type of heuristic algorithms used to solve hard optimization problems where the problem consists of a number of small basis units analogous to genes in a biological system.

This type of algorithm is motivated by the Darwin's natural selection theory. In its essence, it says that if one individual is more likely to be considered of merit according to a given current environment, it is more likely that his/her genes will be kept in the reproductive process. On the contrary, if an individual is perceived to be of bad merits, it is less likely that his/her genes will be preserved from generations to generations. As a result, after years of selections, good genes have a higher relative probability being carried over from generations to generations and it should have a larger presence in a given generation. While for a bad gene, it is less likely to be carried over, so it is less likely to be present or have a smaller ratio in a given generation.

Thus when a given optimization problem contains basic units that take the roles of genes, one can assign a merit value to each such gene. This merit function will in general be positively correlated with the objection function one tries to maximize. For instance, if one gene is generally positively correlated with a high value of the object function, then it should be assigned a high merit value. In the next generation/iteration, one should sample it more frequently.

In general, a genetic algorithm can be divided into four steps: initialization, selection, genetic operators and termination. In the four steps, selection and genetic operators are repeated many time.

In the initialization step, a bunch of individuals are generated with random gene configurations. Each of them will have a merit value. This merit value is different from the merit value for genes. In this case, the merit value is assigned on an individual level and it decides how

likely a given individual is selected in the selection step. As before, this merit value for each individual will be decided by the objective function as well. These individuals make up the first generation of the species.

In the selection steps, a subset of the individual from the previous generation is selected to breed the next generation. The probability a given individual is selected depend on its own merit value. Individuals with high merit values can be selected more than once, while individuals with low merit values might not be selected at all.

The next step in this algorithm is called the genetic operators. There are in general two types of genetic operators, known as crossovers and mutations. Both operators closely mimic the DNA replication in the reproductive process. For crossovers, the genes from the two parents are merged in a deterministic or random way such that the child contains part of the genes from each parent. Mutations are different. It is generally a step that prevents the process being stuck in a local optimum. In this step, part of the child's gene is randomly modified or mutated.

The selection step and the genetic operators are repeated many times until a certain termination criterion is satisfied. This termination criterion usually specifies that the process should stop either when a local maximum is reached or when the process has been repeated long enough.

5.8 Solving the inverse problem

Similar inverse problems were previously solved in the context of quantum mirages[95] using optimization methods including simulated annealing[96] and genetic algorithms[97]. Our approach is primarily based on simulated annealing, with the addition of an adaptive updating step. The simulated annealing algorithm is based on the annealing process in statistical physics, where a solid system is first melted at a high enough temperature and the tempera-

ture is then gradually lowered until the system reaches the ground state. During the annealing process, the energy of the system is gradually being decreased as the temperature is being lowered. For optimization problems, one just needs to define the function to be minimized as the effective energy and ”simulate” the annealing process by introducing an effective temperature.

Without loss of generality, we engineer a pattern resembling the alphabetical letter ”H”. This letter ”H” is first put onto a discretized grid as shown in Fig.5.3(A), where a red box represents a high signal measured in the s-SNOM tip while a blue box corresponds to a low signal(the meaning of ”low” and ”high” will become clear later). We define an effective energy function that measures the ”distance” between any wave pattern from the desired wave pattern. One convenient choice is

$$E = \max_{i \in Blue} m_i - \min_{j \in Red} m_j, \tag{5.12}$$

where m_i is the signal measured at the center of the ith box. This energy function will be gradually decreased as one lowers the system ”temperature” and the resulting wave pattern will approach the desired wave pattern during this annealing process.

The system is initialized with a random configuration and a high enough temperature T such that all possible configurations are equally likely. Simulated annealing is then implemented in an updating step and an acceptance step. In the updating step, a random change to the current configuration is proposed, while in the acceptance step, the change in energy ΔE is calculated and the current configuration is replaced by the proposed configuration with the following probability:

$$P(\Delta E) = \begin{cases} 1, & \Delta E \leqslant 0 \\ exp(-\Delta E/T), & \Delta E > 0 \end{cases}. \tag{5.13}$$

This is the Metropolis-Hastings algorithm[98] and it leads to the Boltzmann distribution in equi-

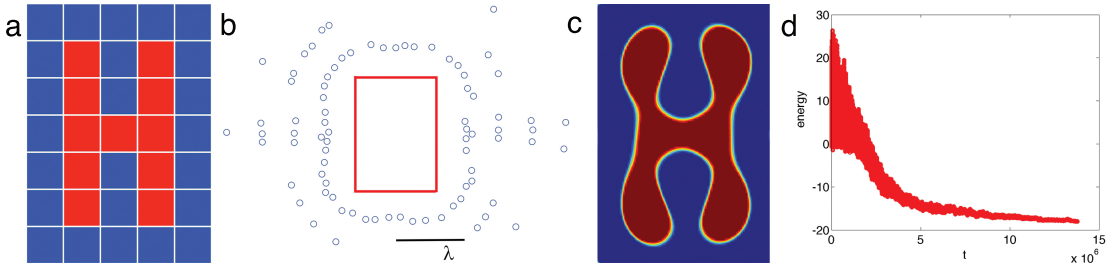

Figure 5.3: Plasmonic Hologram. **(A)** The desired wave pattern discretized on a grid with grid spacing $\lambda/4$. **(B)**The configuration of the nanoparticles that gives rise to the wave pattern "H". The black reference bar denotes one wavelength of the incident light. **(C)** The wave pattern as measured by the s-SNOM tip when the nanoparticles are arranged as in (B). This whole region corresponds to the region inside the red rectangle in (B). **(D)** The energy function at each time step.

librium. After a target number of changes are accepted for a fixed temperature, the system is cooled down to a lower temperature[80] and the whole process repeats until a target temperature is reached or when no more changes are accepted.

We slightly modify the updating step to incorporate information of low energy configurations that have appeared before the current time step. We divide the position space into a finite number of regions and each region is associated with a frequency weight that determines how likely it is going to contribute a nanoparticle. Initially, all regions are equally weighted. At every time step, each nanoparticle is assigned a frequency weight equal to the value of a merit function, and this frequency weight is added to the region that nanoparticle belongs to. After a few time steps, this builds up a frequency profile in space where more frequent regions are more likely to contribute nanoparticles that yield low energy configurations. If one imagines each region as a gene, the merit function then measures the quality of that gene and more frequent regions correspond to genes with higher qualities. We use the following merit function for our simulations:

$$f(E) = \begin{cases} (1-\frac{E}{2T})^2, & 0 < E < 2T \\ min(exp(-E/T), exp(5)), & E \leqslant 0 \\ 0, & otherwise \end{cases} . \tag{5.14}$$

In the updating step, we randomly pick a nanoparticle from the current configuration. With probability p, we uniformly generate a new random position for it and with the other 1-p probability, we generate a random position in its neighboring regions according to the current frequency profile, favoring ”genes” with higher qualities.

The resulting configuration of the nanoparticles is shown in Fig.5.3(B) and the corresponding signal pattern is plotted in Fig.5.3(C). When plotting Fig.5.3(C), our algorithm finds a threshold signal value m_c such that when the signal measured at the current tip position is smaller than m_c, it is colored blue(low signal), and red(high signal) otherwise. In these simulations, we use 80 gold nanospheres with diameter D=20nm and resonance wavelength $\lambda = 563nm$[20]. The s-SNOM tip is assumed to be polarized in the z direction. To guarantee the generality of the algorithm, we do not impose any symmetry constraint on the configuration of the nanoparticles in the simulations. Interestingly, a quasi-symmetric pattern emerges in Fig.5.3(B).

A plot of the energy at each time step is shown in Fig.5.3(D). As one can see, the energy on average decreases over time with diminishing fluctuations. The average energy for random configurations in this case is 11.9(arbitrary unit) with standard deviation 4.78. The final energy for the configuration in Fig.5.3(C) is -18.03, which is more than six standard deviations below the average.

In the above case, the s-SNOM tip acts both as a signal probe and an illumination source. Since the source itself is moving, the signal being read out is not the scattering wave pattern itself, but a result of the interference between the incident wave and the backscattered wave at each tip position[81,82,88,90,86,87].

In Fig.5.4, we show the frequency profile that builds up at the end of the simulations. It clearly shares certain similarity to Fig.5.3(B). This is comforting in the sense that the result is converging. However, it by no means guarantee that we have found a global minimum. Also

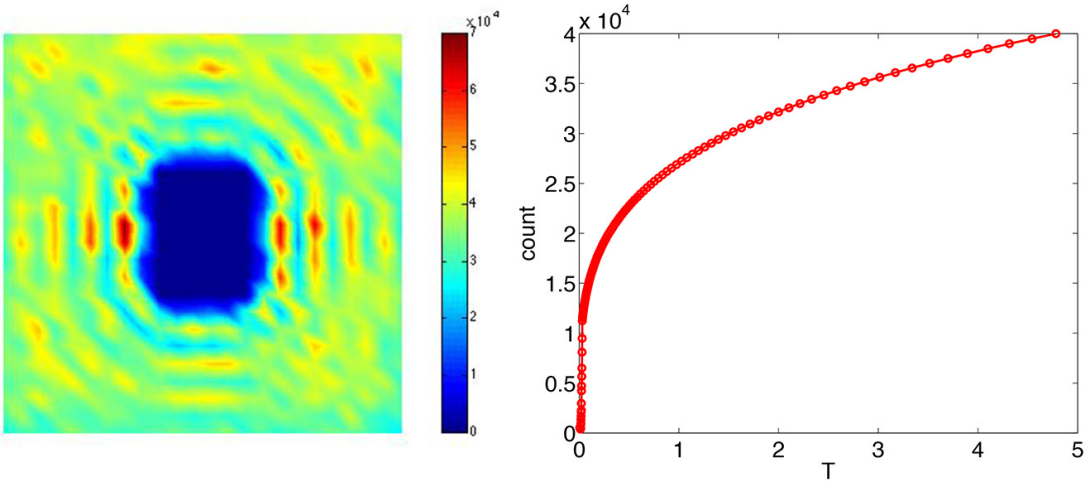

Figure 5.4: Frequency profile and temperature dependence of the number of changes accepted. Left: The frequency profile that builds up at the end of the simulations. Clearly, it shares certain similarity to the profile in Fig.5.3(B). Right: The number of changes accepted at each temperature.

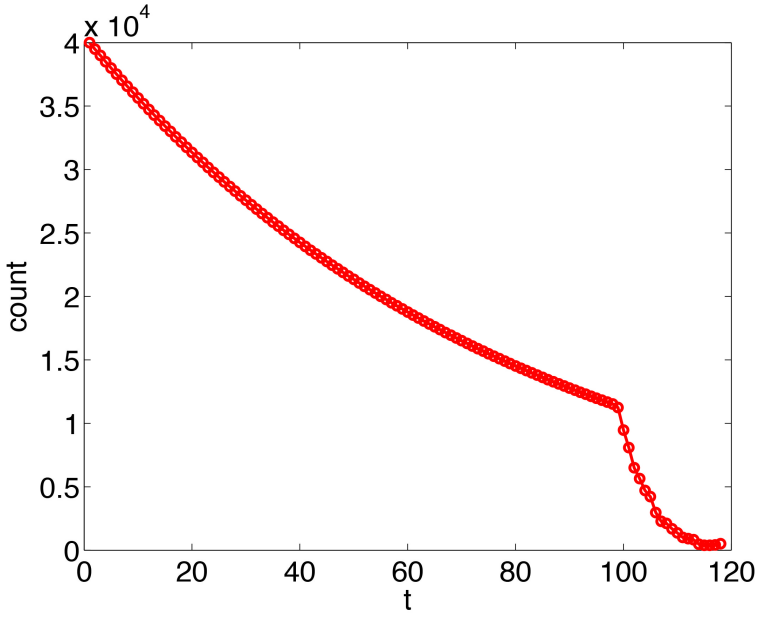

Figure 5.5: Number of changes accepted at each "thermal equilibrium". The number of changes accepted at each thermal equilibrium. Here t labels the temperature and a kink is clearly visible. In the simulations, we decrease the temperature either when $4000(\frac{T}{T_0})^{\frac{1}{4}}$ changes have been accepted, where T_0 is the initial temperature, or when 80000 changes have been proposed.

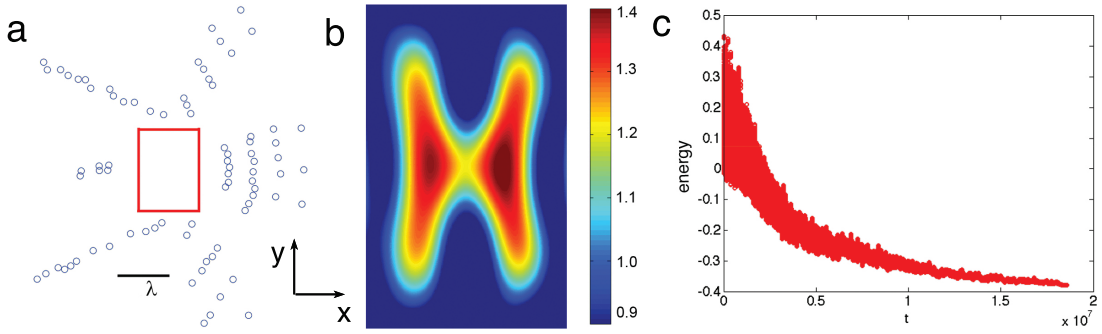

Figure 5.6: Engineering the illumination pattern. **(A)** The configuration of the nanoparticles that gives rise to a scattering wave pattern resembling the letter "H". We have imposed a minimum spacing constraint on each nanoparticle's position in the algorithm, so no two nanoparticles are overlapping. The blue circles in this plot is significantly enlarged for better visualization.**(B)** The scattering wave pattern($|E_z|/|E_0|$) that is generated given the configuration in (A) and a plane wave incident from the x direction. **(C)** The energy function at each time step.

shown in Fig.5.4 is the number of changes accepted at each temperature. The data here is a little bit more misleading because we require a certain number of changes to be accepted at each temperature, so the total number of proposals is not the same for each temperature. However, a kink still develops at lower temperature as shown in the figure below.

A possibly more interesting problem is to engineer the illumination pattern itself, that is, to engineer the scattering wave pattern given a fixed incident wave. Subwavelength control of this illumination pattern can be relevant for a wide range of scenarios, including but not limited to the design of more flexible platforms for probing many body interaction[40,99,100,101,102].

Consider an incident plane wave of the form $\vec{E}_0(\vec{r}) = E_0 e^{ikx}\hat{z}$. The same approach applies with modified incident conditions. The results are shown in Fig.5.6. In this case, the average energy(normalized by the magnitude of the incident wave) for completely random configurations is 0.2 with standard deviation 0.06. The final energy for the pattern in Fig.5.7(B) is -0.38, which is 9.7 standard deviations below the average. Moreover, the shaped electromagnetic fields are spatially separated from the scatterers themselves. Thus, one can potentially employ this approach to build optical traps for manipulating ultracold atoms' interactions without introducing undesired sources of decoherence. The corresponding frequency profile

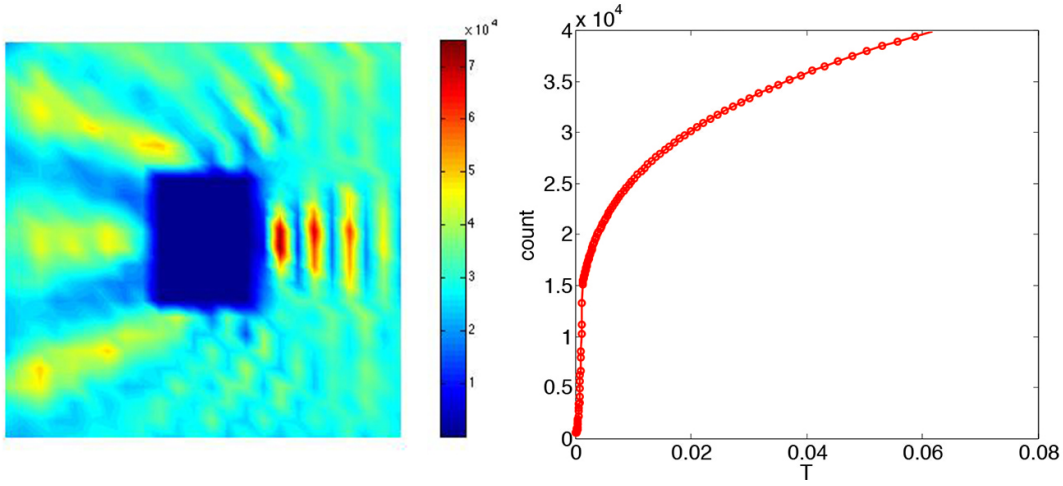

Figure 5.7: Frequency profile and temperature dependence of the number of changes accepted for the illumination pattern case. Left: The frequency profile that builds up at the end of the simulations. Clearly, it shares certain similarity to the profile in Fig.5.6(B). Right: The number of changes accepted at each temperature.

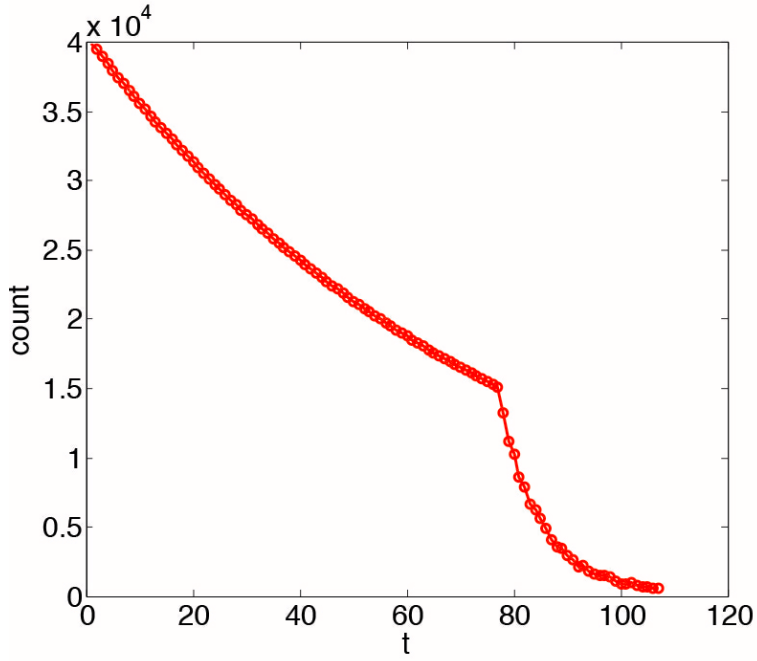

Figure 5.8: Number of changes accepted at each "thermal equilibrium" for the illumination pattern case.The number of changes accepted at each thermal equilibrium. Here t labels the temperature and a kink is clearly visible, possibly indicating a phase transition.

and accepted changes are shown in Fig.5.7 and Fig.5.8.

5.9 Summary

In summary, we have shown that one can gain a considerable amount of control over electromagnetic fields by solving the inverse scattering problem. In our examples, we use metallic nanoparticles as the scatterers of light. By engineering their shapes and sizes, this design could be made effective in the whole visible spectrum[103]. To extend this approach to other spectra, one needs to use different types of scatterers. For instance, one can replace the metallic nanoparticles with graphene nanostructures/microstructures for the infrared spectrum[24,25]. Our design does not require spatial engineering of a material's electromagnetic properties and has relevance to developing more flexible platforms for probing light matter interaction or many body effects.

A

Random Potential Generation

There are two major kinds of disorders in the experimental system[15]. The first one is from the charged donor atoms in the donor layer, whose distance to the 2DEGs can be controlled experimentally. This distance is related to the correlation length of the random potential. These randomly placed donor atoms contribute to a weakly correlated random potential which barely does any backscattering.

The second one is from the impurity atoms, which can lie very close to the 2DEGs, resulting in a sharp potential that strongly backscatters.

It is established both theoretically[9,15] and experimentally[11] that the random potential from the donors is mainly responsible for branching while the impurity potential is responsible for fringing, which is another striking observation of the experiment[2,10]. For the experimental results[11] we are interested in, fringing is significantly suppressed by using very high mobility samples[11]. Thus, we shall focus only on the random potential due to donor atoms.

One simple model that takes into account screening yields the following result[1] for the potential due to one charged donor atom:

$$V(r) = -q\frac{\hbar^2}{2m}\frac{d}{(r^2+d^2)^{3/2}}, \tag{A.1}$$

where q is the charge on the donor atom, d is the donor to 2DEGs distance and r is the distance measured in-plane.

Given an estimate of donor density in the experiment, one can then generate a random potential using the above model. However, in order to match the mobility of the experimental samples, one needs to adjust the standard deviation of the strength of the random potential. As estimated by both us and the experimental group in their supplementary material[11], the resulting random potential has a standard deviation of $8\%E_F$.

To characterize the legnth scale in the random potential, we define a correlation length through the autocorrelation function of the potential. Given a random potential $V(\vec{r})$, the autocorrelation function is defined to be $< V(\vec{r}' - \vec{r})V(\vec{r}') >$, where the bracket means sample or ensemble average over $\vec{r}'$. It is approximately true that the autocorrelation function has the following scaling relation:

$$< V(\vec{r}' - \vec{r})V(\vec{r}') > \sim e^{-\frac{r^2}{l_c^2}}, \tag{A.2}$$

where l_c is defined to be the correlation length. For our sample, the correlation length is esti-

mated to be $l_c = 0.9\lambda_F$.

B

Classical Simulation Methods

The classical simulations are based on the Newton's equations and the only trick lies in how to choose the initial conditions.

For the classical simulations, we use Monte Carlo method to generate 10^7 classical particles with the following initial distribution in position and momentum, which is a transformation of the initial quantum wav packet using the Wigner's quasi-probability distribution:

$$P(x,p_x) = \frac{1}{\pi\sigma_{p_x}\sigma_x} exp(-\frac{p_x^2}{\sigma_{p_x}^2} - \frac{x^2}{\sigma_x^2}), \tag{B.1}$$

where $\sigma_{p_x}^2 = m\hbar\omega(y_0)$ and $\sigma_x^2 = \hbar/m\omega(y_0)$.

The advantages of this transformation, as mentioned in the main text, are: (1) it properly accounts for the uncertainty principle and (2) it reproduces the quantum marginal position and momentum distribution.

Additionally, to keep the matters simple, the energy of each particle is chosen to be fixed at E_F. This is guaranteed by eliminating trajectories with larger energy and boosting those with less in p_y.

The time evolution of those particles is governed by the classical equation of motion and the results are presented by the method of Poincare's surface of section at fixed y. Essentially, the idea behind Poincare's surface of section is to reduce the dimension of phase space to two by fixing y. Each classical particle can be represented in the four dimensional phase space by (x, y, p_x, p_y). Since we are dealing with classical particles of fixed energy, the constraint imposed by the conservation of energy reduces the phase space structure to three dimensions. The two dimension plots shown in the main text are done by further fixing y, which decreases the degree of freedom by one. If one is interested in recovering the full three dimensional structure in phase space, one can plot many two-dimensional plots at different y and combine them. For our purpose, we are only interested in the saturated overlap produced by the QPC. Therefore, we only show plots at a y that is sufficiently far away from the QPC, where the overlap has saturated to a constant value.

For the classical flux simulations shown in the main text, we use the same Monte Carlo methods to sample the distribution given by (B.1) and calculate the total flux through a given point by counting the total number of classical particles that pass through a small neighborhood of that point with weights determined by their momenta in the y direction.

From a numerical point of view, the most tricky part of this classical simulation is to guarantee energy conservation. The Newton's equations, in its continuous form, guarantees energy

conservation, but when one converts it to the finite difference form, energy is no longer conserved. Thus, one needs to reenforce energy conservation at each time step.

References

[1] M.A. Topinka, B. J. LeRoy, S. E. J. Shaw, E. J. Heller, R. M. Westervelt, K. D. Maranowski and A. C. Gossard, Science **289**, 2323 (2000).

[2] M. A. Topinka, B. J. LeRoy, R. M. Westervelt, S. E. J. Shaw, R. Fleischmann, E. J. Heller, K. D. Maranowski and A. C. Gossard, Nature **410**, 183-186 (2001).

[3] Bo Liu and Eric J. Heller, Phys. Rev. Lett. **111**, 236804(2013).

[4] R. Hohmann, U. Kuhl, H.-J. Stockmann, L. Kaplan and E. J. Heller, Phys. Rev. Lett **104**, 093901 (2010).

[5] M. A. Wolfson and S. Tomsovic, J. Acoust. Soc. Am. **109**, 2693 (2001).

[6] K. E. Aidala, R. E. Parrott, T. Kramer, E. J. Heller, R. M. Westervelt, M. P. Hanson and A. C. Gossard, Nat. Phys. **3**, 464 (2007).

[7] D. Maryenko, F. Ospald, K. V. Klitzing, J. H. Smet, J. J. Metzger, R. Fleischmann, T. Geisel and V. Umansky, Phys. Rev. B **85**, 195329 (2012).

[8] E. J. Heller, L. Kaplan, and A. Dahlen, J. Geophys.Res. **113**, C09023(2008).

[9] M. P. Jura, M. A. Topinka, L. Urban, A. Yazdani, H. Shtrikman, L. N. Pffer, K. W. West and D. Goldhaber-Gordon, Nat. Phys. **3**, 841-845 (2007).

[10] E. J. Heller, K. E. Aidala, B. J. LeRoy, A. C. Bleszynski, A. Kalben, R. M. Westervelt, K. D. Maranowski and A. C. Gossard, Nano Lett. **5**, 1285 (2005).

[11] S. E. J. Shaw's Thesis, Harvard University, available at http://www.physics.harvard.edu/Thesespdfs/sshaw.pdf (2002).

[12] Adel Abbout, Gabriel Lemarie and Jean-Louis Pichard, Phys. Rev. Lett **106**, 156810 (2011).

[13] E. J. Heller, J. Chem. Phys. **65**, 1289 (1976).

[14] L. Kaplan, Phys. Rev. Lett **89**, 184103 (2002).

[15] E. J. Heller and S. E. J. Shaw, International Journal of Modern Physics B **17**, 3977 (2003).

[16] J. J. Metzger, R. Fleischmann and T. Geisel, Phys. Rev. Lett **105**, 020601 (2010).

[17] Jakob J. Metzger, PhD thesis, University of Gottingen, available at http://www.chaos.gwdg.de/publications/phdthesisreference.2011-01-13.8079857018

[18] Mark Topinka, PhD thesis, Harvard University, available at http://meso.seas.harvard.edu/theses.html

[19] S.A. Maier, Plasmonics: Fundamentals and Applications, Springer: New York,2007.

[20] C. Sonnichsen, T. Franzl, T. Wilk, G. von Plessen, J. Feldmann, O. Wilson, P. Mulvaney, Phys. Rev. Lett. **88**, 077402(2002).

[21] S. Eustis, M. A. El-Sayed, Chem. Soc. Rev.**35**, 209(2006).

[22] N. J. Halas, S. Lal, W. Chang, S. Link, P. Nordlander, Chem. Rev. **111**, 3913(2011).

[23] E. Prodan, C. Radloff, N. J. Halas, P. Nordlander, Science **302**, 419(2003).

[24] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H. A. Bechtel, X. Liang, A. Zettl, Y. R. Shen, F. Wang, Nature Nanotechnology **6**, 630(2011).

[25] H. Yan, X. Li, B. Chandra, G. Tulevski, Y. Wu, M. Freitag, W. Zhu, P. Avouris, F. Xia, Nature Nanotechnology **7**, 330(2012).

[26] J. A. Schuller, E. S. Barnard, W. Cai, Y. C. Jun, J. S. White, M. L. Brongersma, Nature Materials **9**, 193(2010).

[27] A. G. Brolo, Nature Photonics **6**, 709 (2012).

[28] R. Shelby, D. R. Smith, S. Schultz, Science **292**, 77 (2001).

[29] M. L. Juan, M. Righini, R. Quidant, Nature Photonics **5**, 349 (2011).

[30] H. A. Atwater, A. Polman, Nature Materials **9**, 205 (2010).

[31] E. M. Hicks, S. Zou, G. C. Schatz, K. G. Spears, R. P. Van Duyne, L. Gunnarsson, T. Rindzevicius, B. Kasemo, and M. Kall, Nano Lett. **5** 1065(2005).

[32] J. Zuloaga, E. Prodan, P. Nordlander, Nano Lett. **9**, 887(2009).

[33] A. Aubry, D. Y. Lei, A. I. Fernandez-Dominguez, Y. Sonnefraud, S. A. Maier, J. B. Pendry, Nano Lett. **10**, 2574(2010).

[34] A. Aubry, D. Y. Lei, S. A. Maier, J. B. Pendry, Phys. Rev. Lett. **105**, 233901(2010).

[35] S. Kim, J. Jin, Y. J. Kim, I. Y. Park, Y. Kim, S. W. Kim, Nature **453**, 757(2008).

[36] P. Nordlander, C. Oubre, E. Prodan, K. Li, M. I. Stockman, Nano Lett. **4**, 899(2004).

[37] E. J. Heller, Phys. Rev. Lett. **77**, 4122(1996).

[38] S. Thongrattanasiri, F. H. L. Koppens, F. J. Garcia de Abajo, Phys. Rev. Lett. **108**, 047401(2012).

[39] D. V. van Coevorden, R. Sprik, A. Tip, A. Lagendijk, Phys. Rev. Lett. **77**, 2412(1996).

[40] K. J. Savage, M. M. Hawkeye, R. Esteban, A. G. Borisov, J. Aizpurua, J. J. Baumberg, Nature **491**, 574(2012).

[41] F. Javier Garcia de Abajo, Rev. Mod. Phys. **79**, 1267(2007).

[42] Giuseppe Grosso and Giuseppe Pastori Parravicini, Solid State Physics, Academic Press, 2000.

[43] S. Y. Park, D. Stroud, Phys. Rev. B **69**, 125418(2004).

[44] J. D. Jackson, Classical Electrodynamics, John Wiley & Sons: New York, 1998.

[45] M. Chu, V. Myroshnychenko, C. H. Chen, J. Deng, C. Mou, F. Javier Garcia de Abajo, Nano Lett. **9**,399(2009).

[46] S. Nie and S. R. Emory, Science **275**, 1102(1997).

[47] S. A. Maier, Nature Materials **8**, 699(2009).

[48] P. Nordlander, ACS Nano **3**, 488(2009).

[49] N. Liu, L. Langguth, T. Weiss, J. Kastel, M. Fleischhauer, T. Pfau, H. Giessen, Nature Materials **8**, 758(2009).

[50] N. Liu, M. Hentschel, T. Weiss, A. P. Alivisatos, H. Giessen, Science **332**, 1407(2011).

[51] A. Wokaun, J. P. Gordon, P. F. Liao, Phys. Rev. Lett. **48**, 957(1982).

[52] S. A. Maier, P. G. Kik, H. A. Atwater, S. Meltzer, E. Harel, B. E. Koel, A. A. G. Requicha, Nature Materials **2**, 229(2003).

[53] Mark L. Brongersma, John W. Hartman, and Harry A. Atwater, Phys. Rev. B **62**, 16356(2000).

[54] Stefan A. Maier, Pieter G. Kik, Harry A. Atwater, Sheffer Meltzer, Elad Harel, Bruce E. Koel, and Ari AG Requicha, Nature Materials **2**, 229(2003).

[55] Stefan A. Maier, Pieter G. Kik, and Harry A. Atwater, Applied Physics Letters **81**, 1714(2002).

[56] Dmitri K. Gramotnev, and Sergey I. Bozhevolnyi, Nature Photonics **4** , 83(2010).

[57] Ekmel Ozbay, Science **311**, 189(2006).

[58] Piotr J. Flatau, and B. T. Draine, J. Opt. Soc. Am. A **11**, 1491(1994).

[59] Harry A. Atwater, Scientific American **296**, 56(2007).

[60] Esteban Moreno, Sergio G. Rodrigo, Sergey I. Bozhevolnyi, L. Martin-Moreno, and F. J. Garcia-Vidal, Phys. Rev. Lett. **100**, 023901(2008).

[61] Sergey I. Bozhevolnyi, Valentyn S. Volkov, Eloise Devaux, Jean-Yves Laluet, and Thomas W. Ebbesen, Nature **440**, 508 (2006).

[62] Kanglin Wang, and Daniel M. Mittleman, Nature **432**,376(2004).

[63] P. Andrew, and W. L. Barnes, Science **306**, 1002(2004).

[64] M. Quinten, A. Leitner, J. R. Krenn, and F. R. Aussenegg , Optics Letters **23**, 1331(1998).

[65] Mark L. Brongersma, John W. Hartman, and Harry A. Atwater, Phys. Rev. B **62**, 16356(2000).

[66] Stefan A. Maier, Pieter G. Kik, Harry A. Atwater, Sheffer Meltzer, Elad Harel, Bruce E. Koel, and Ari AG Requicha, Nature Materials **2**, 229(2003).

[67] Stefan A. Maier, Pieter G. Kik, and Harry A. Atwater, Applied Physics Letters **81**, 1714(2002).

[68] L. Tang, S. E. Kocabas, S. Latif, A. K. Okyay, Dany-Sebastien Ly-Gagnon, K. C. Saraswat, D. AB Miller, Nature Photonics **2**, 226(2008).

[69] D. E. Chang, J. D. Thompson, H. Park, V. Vuletic, A. S. Zibrov, P. Zoller, M. D. Lukin, Phys. Rev. Lett. **103**, 123004(2009).

[70] D. R. Smith, J. B. Pendry, and M. C. K. Wiltshire, Science **305**, 788(2004).

[71] D. Schurig, J. J. Mock, B. J. Justice, Steven A. Cummer, J. B. Pendry, A. F. Starr, and D. R. Smith, Science **314**, 977(2006).

[72] N. I. Landy, S. Sajuyigbe, J. J. Mock, D. R. Smith, and W. J. Padilla, Phys. Rev. Lett. **100**, 207402(2008).

[73] John B. Pendry, David Schurig, and David R. Smith, Science **312**, 1780(2006).

[74] J. B. Pendry, A. Aubry, D. R. Smith, and S. A. Maier, Science **337**, 549(2012).

[75] R. Liu, C. Ji, J. J. Mock, J. Y. Chin, T. J. Cui, and D. R. Smith, Science **323**, 366(2009).

[76] J. B. Pendry, Phys. Rev. Lett. **85**, 3966(2000).

[77] Y. Lai, J. Ng, H. Chen, D. Han, J. Xiao, Z. Zhang, C. T. Chan, Phys. Rev. Lett. **102**, 253902 (2009).

[78] Suenne Kim, Farbod Shafiei, Daniel Ratchford, and Xiaoqin Li, Nanotechnology **22**, 115301(2011).

[79] A. Silva, F. Monticone, G. Castaldi, V. Galdi, A. Alan, and N. Engheta, Science **343**,160(2014).

[80] S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, Science **220**, 4598(1983).

[81] G. A. Fiete, E. J. Heller, Rev. Mod. Phys. **75**, 933(2003).

[82] C R. Moon, L. S. Mattos, B. K. Foster, G. Zeltzer, H. C. Manoharan, Nature Nanotechnology **4**, 167(2009).

[83] Michael F. Crommie, Christopher P. Lutz, and Donald M. Eigler, Science **262**, 218 (1993).

[84] E. J. Heller, M. F. Crommie, C. P. Lutz and D. M. Eigler, Nature **369** , 464 (1994).

[85] F. Zenhausern, Y. Martin, and H. K. Wickramasinghe, Science **269**, 1083 (1995).

[86] G. Colas des Francs, C. Girard, J. C. Weeber, C. Chicane, T. David, A. Dereux, D. Peyrade, Phys. Rev. Lett. **86**, 4950(2001).

[87] C. Chicane, T. David, R. Quidant, J. C. Weeber, Y. Lacroute, E. Bourillot, A. Dereux, G. Colas des Francs, C. Girard, Phys. Rev. Lett. **88**, 097402(2002).

[88] F. Keilmann, R. Hillenbrand, Phil. Trans. R. Soc. Lond. A **362**, 787(2004).

[89] R. Hillenbrand, F. Keilmann, Phys. Rev. Lett. **85**, 3029(2000).

[90] Z. Fei, A. S. Rodin,G. O. Andreev,W. Bao, A. S. McLeod,M. Wagner, L. M. Zhang, Z. Zhao, M. Thiemens, G. Dominguez, M. M. Fogler, A. H. Castro Neto, C. N. Lau, F. Keilmann , D. N. Basov, Nature **487**, 82(2012).

[91] J. Chen et. al., Nature **487**, 77(2012).

[92] M. L. Juan, M. Righini, R. Quidant, Nature Photonics **5**, 349 (2011).

[93] A.N. Grigorenko, N. W. Roberts, M. R. Dickson, Y. Zhang, Nature Photonics **2**, 365(2008).

[94] C. Girard, A. Dereux, O. J. F. Martin, and M. Devel, Phys. Rev. B **52**, 2889 (1995).

[95] A. A. Correa, F. A. Reboredo, C. A. Balseiro, Phys. Rev. B, **71**, 035418(2005).

[96] Peter J. M. Van Laarhoven, and Emile H. L. Aarts, Simulated annealing, Springer Netherlands, 1987.

[97] David E. Goldberg, Genetic algorithms in search, optimization, and machine learning, Reading Menlo Park: Addison-wesley, 1989.

[98] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, J. Chem. Phys. **21**, 1087(1953).

[99] J. Berthelot et. al., Nature Nanotechnology **9**, 295(2014).

[100] J. Simon, W. S. Bakr,R. Ma, M. E. Tai, P. M. Preiss, M. Greiner, Nature **472**, 307(2011).

[101] M. Gullans, T. G. Tiecke, D. E. Chang, J. Feist, J. D. Thompson, J. I. Cirac, P. Zoller, M. D. Lukin, Phys. Rev. Lett. **109**, 235309(2012).

[102] C. Stehle, H. Bender, C. Zimmermann, D. Kern, M. Fleischer, S. Slama, Nature Photonics **5**, 494(2011).

[103] K. L. Kelly, E. Coronado, L. Zhao, G. C. Schatz, J. Phys. Chem. B **107**, 668(2003).

[104] H. Yan, X. Li, B. Chandra, G. Tulevski, Y. Wu, M. Freitag, W. Zhu, P. Avouris, F. Xia, Nature Nanotechnology **7**, 330(2012).